On Higher-Order Fourier Analysis over Non-Prime Fields

Arnab Bhattacharyya[1], Abhishek Bhowmick[2], and Chetan Gupta[3]

1 **Department of Computer Science & Automation, Indian Institute of Science, India**
arnabb@csa.iisc.ernet.in

2 **Department of Computer Science, The University of Texas at Austin, USA**
bhowmick@cs.utexas.edu

3 **Department of Computer Science & Automation, Indian Institute of Science, India**
chetan.gupta@csa.iisc.ernet.in

Abstract

The celebrated Weil bound for character sums says that for any low-degree polynomial P and any additive character χ, either $\chi(P)$ is a constant function or it is distributed close to uniform. The goal of higher-order Fourier analysis is to understand the connection between the algebraic and analytic properties of polynomials (and functions, generally) at a more detailed level. For instance, what is the tradeoff between the equidistribution of $\chi(P)$ and its "structure"?

Previously, most of the work in this area was over fields of prime order. We extend the tools of higher-order Fourier analysis to analyze functions over general finite fields. Let $\mathbb{K}$ be a field extension of a prime finite field $\mathbb{F}_p$. Our technical results are:

1. If $P : \mathbb{K}^n \to \mathbb{K}$ is a polynomial of degree $\leqslant d$, and $\mathbb{E}[\chi(P(x))] > |\mathbb{K}|^{-s}$ for some $s > 0$ and non-trivial additive character χ, then P is a function of $O_{d,s}(1)$ many *non-classical polynomials* of weight degree $< d$. The definition of non-classical polynomials over non-prime fields is one of the contributions of this work.
2. Suppose $\mathbb{K}$ and $\mathbb{F}$ are of bounded order, and let H be an affine subspace of $\mathbb{K}^n$. Then, if $P : \mathbb{K}^n \to \mathbb{K}$ is a polynomial of degree d that is sufficiently regular, then $(P(x) : x \in H)$ is distributed almost as uniformly as possible subject to constraints imposed by the degree of P. Such a theorem was previously known for H an affine subspace over a prime field.

The tools of higher-order Fourier analysis have found use in different areas of computer science, including list decoding, algorithmic decomposition and testing. Using our new results, we revisit some of these areas.

(i) For any fixed finite field $\mathbb{K}$, we show that the list decoding radius of the generalized Reed Muller code over $\mathbb{K}$ equals the minimum distance of the code.

(ii) For any fixed finite field $\mathbb{K}$, we give a polynomial time algorithm to decide whether a given polynomial $P : \mathbb{K}^n \to \mathbb{K}$ can be decomposed as a particular composition of lesser degree polynomials.

(iii) For any fixed finite field $\mathbb{K}$, we prove that all locally characterized affine-invariant properties of functions $f : \mathbb{K}^n \to \mathbb{K}$ are testable with one-sided error.

1998 ACM Subject Classification G.2 Discrete Mathematics, D.2.5 Testing and Debugging

Keywords and phrases finite fields, higher order fourier analysis, coding theory, property testing

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2016.23

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2016).
Editors: Klaus Jansen, Claire Matthieu, José D. P. Rolim, and Chris Umans; Article No. 23; pp. 23:1–23:29

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

In this work, we provide new results about polynomials over finite fields, relating their algebraic structure to the distribution of their output. We then apply them to improve previous work on list-decoding bounds for the Reed-Muller code, testability of affine-invariant properties, and algorithms for polynomial decomposition.

1.1 Structure versus Randomness for Polynomials over Finite Fields

In many areas of mathematics, there is a remarkable phenomenon where natural objects are either close to random or have a high degree of structure. A prime example of this is the Weil bound for character sums [49], a deep result in algebraic geometry.

Let $\mathbb{F}$ be a finite field of prime order p, and let $\mathbb{K}$ be a finite field extension of $\mathbb{F}$. Let $P : \mathbb{K} \to \mathbb{K}$ be a univariate polynomial of degree $\leqslant |\mathbb{K}|^{1/2-\delta}$ for some $\delta > 0$. If we let $\chi : \mathbb{K} \to \mathbb{C}$ denote a non-trivial additive character, then according to Weil's bound, either $\chi(P(x))$ is constant or else, $\chi(P(x))$ is distributed close to uniform in the sense that $|\mathbb{E}[\chi(P(x))]| \leqslant |\mathbb{K}|^{-\delta}$. Deligne later [17] proved the same statement for multivariate polynomials $P : \mathbb{K}^n \to \mathbb{K}$.

Higher-order Fourier analysis [46] is a recent generalization of some aspects of Fourier analysis. Over finite fields, one of the main components of the theory is a detailed study of the interplay between the algebraic structure and analytic properties of polynomials. Consider the case when $\mathbb{K} = \mathbb{F}_p$ is a prime field and p is small, for example 2. Weil's bound does not apply for $d > \sqrt{p}$. However, in the context of higher-order Fourier analysis, Kaufmann and Lovett [37] (extending previous work by Green and Tao [30]) showed that if $P : \mathbb{F}_p^n \to \mathbb{F}_p$ is a polynomial of degree d, then for any non-trivial additive character χ, either $|\mathbb{E}[\chi(P(x))]| < \varepsilon$ or else, P is a function of a $O_{\varepsilon,d,p}(1)$ polynomials of strictly lesser degree. The regime here is different from that of Weil's bound (large n versus large $|\mathbb{K}|$) but the result is similar in spirit.

These recent developments have spurred a deeper look at the dichotomy between randomness and structure in polynomials over finite fields. For instance, instead of using the bias, $|\mathbb{E}[\chi(P(x))]|$, as a measure of randomness, one can look at how well P is equidistributed on affine subspaces. This gives rise to the *Gowers norm* [28, 29], a central notion in higher-order Fourier analysis. Low Gowers norm is a much stronger notion of pseudorandomness than low bias. Green and Tao [30] showed the *Gowers inverse theorem*, which states that if $P : \mathbb{F}_p^n \to \mathbb{F}_p$ is a polynomial of degree $d < p$, then either its Gowers norm of order d is smaller than ε or it is a function of $O_{\varepsilon,d,p}(1)$ polynomials of degree $< d$. When $d \geqslant p$, this dichotomy fails to be true [40, 30]; Tao and Ziegler [47] showed that if the Gowers norm is not small, then the polynomial is a function of a bounded number of *non-classical polynomials*, functions mapping to the torus that locally look like low-degree polynomials.

All of these results focused on finite fields of prime order. In general fields, the situation is more complicated for several reasons. Firstly, the additive characters over $\mathbb{F}$ are simply exponential functions $\chi_a(x) = e^{2\pi i a x/p}$ for $a \neq 0$ which are bijective, while over general fields $\mathbb{K}$, the additive characters are $\chi_a(x) = e^{2\pi i \mathrm{Tr}(ax)/p}$ for $a \neq 0$ where the trace $\mathrm{Tr} : \mathbb{K} \to \mathbb{F}$ is a linear map with a possibly large kernel. Secondly, if $\mathbb{K}$ is of dimension at least 2 over $\mathbb{F}$, polynomials like x^p have degree p but vanish after taking only 2 derivatives. Define the *weight degree* of a polynomial $P : \mathbb{K}^n \to \mathbb{K}$ to be the minimum number of derivatives before P becomes identically a constant. If a variable has individual degree more than p in a monomial in P, then the weight degree of P may not equal the total degree. Thirdly, while $\mathbb{K}$ can often be profitably viewed as simply a vector space over $\mathbb{F}$, affine subspaces in $\mathbb{K}^n$ are not necessarily vector spaces over $\mathbb{F}$.

In this work, we prove the first structure-versus-randomness dichotomy result for polynomials over general fields in the higher-order Fourier analysis setting. We show that if $P : \mathbb{K}^n \to \mathbb{K}$ is a polynomial of total degree d, then either its Gowers norm of order d is less than $|\mathbb{K}|^{-s}$ (and hence, so is the bias of any additive character of P), or else, P is a function of $O_{d,s}(1)$ many *non-classical polynomials* over $\mathbb{K}^n$ of weight degree less than d. Our definition of non-classical polynomials uses the multiplicative structure of $\mathbb{K}$. Also, note that unlike the results quoted above, our result is non-trivial when $|\mathbb{K}|$ is not constant. For constant d, our result continues to give information about the structure of the polynomial even when its bias is less than $|\mathbb{K}|^{-1/2}$, the limit of Weil's bound. To make our bounds hold in this regime, we use some results recently given by Bhowmick and Lovett [13] where they mainly[1] studied higher-order Fourier analysis over growing prime fields.

Our next dichotomy result holds for constant sized $\mathbb{K}$. Let c be a positive integer bound, and let $P : \mathbb{K}^n \to \mathbb{K}$ be a polynomial of weight degree d. Then, either for some $\alpha \in \mathbb{K}$, $\mathrm{Tr}(\alpha P)$ is expressible in terms of polynomials of lower degree, or the distribution of P on random affine subspaces of $\mathbb{K}^n$ of dimension c is as uniform as possible subject to the constraints imposed by the weight degree of P. For instance, if the weight degree of P is 1, then $P(x+y+z) - P(x+y) - P(x+z) + P(x) = 0$ for all $x, y, z \in \mathbb{K}^n$ and hence, this constraint must be satisfied by the evaluations of P on any affine subspace. The second possibility in the dichotomy is that modulo such constraints, the value of P is almost unconstrained on subspaces of dimension c.

1.2 Applications

In this section, we describe three different problems involving a finite field $\mathbb{K}$, which previously had been solved only when $|\mathbb{K}|$ was prime but which we can now solve for arbitrary finite $\mathbb{K}$.

Throughout, let $\mathbb{F}$ be a fixed prime order field, and let $\mathbb{K}$ be a finite field that extends $\mathbb{F}$. Let $q = |\mathbb{K}|$, $p = |\mathbb{F}|$ and $q = p^r$ for $r > 0$.

1.2.1 List-decoding Reed-Muller codes

The notion of *list decoding* was introduced by Elias [18] and Wozencraft [50] to decode *error correcting codes* beyond half the minimum distance. The goal of a list decoding algorithm is to produce all the codewords within a specified distance from the received word. At the same time one has to find the right radius for which the number of such codewords is small, otherwise there is no hope for the algorithm to be efficient. After the seminal results of Goldreich and Levin [20] and Sudan [43] which gave list decoding algorithms for the Hadamard code and the Reed-Solomon code respectively, there has been tremendous progress in designing list decodable codes. See the survey by Guruswami [34, 33] and Sudan [44].

Reed-Muller codes (RM codes) were discovered by Muller in 1954. Let $d \in \mathbb{N}$. The RM code $\mathrm{RM}_{\mathbb{K}}(n, d)$ is defined as follows. The message space consists of degree $\leqslant d$ polynomials in n variables over $\mathbb{K}$ and the codewords are evaluation of these polynomials on $\mathbb{K}^n$. Let $\delta_q(d)$ denote the normalized distance of $\mathrm{RM}_{\mathbb{K}}(n, d)$. Let $d = a(q-1) + b$ where $0 \leqslant b < q-1$. We have

$$\delta_{\mathbb{K}}(d) = \frac{1}{q^a}\left(1 - \frac{b}{q}\right).$$

[1] As we discuss later, they also study non-prime fields in Section 4.9 but they restrict to the case $d < p$ whereas we focus on small p.

RM codes are one of the most well studied error correcting codes. Many applications in computer science involve low degree polynomials over small fields, namely RM codes. Given a received word $g : \mathbb{K}^n \to \mathbb{K}$ the objective is to output the list of codewords (e.g. low-degree polynomials) that lie within some distance of g. Typically we will be interested in regimes where list size is either independent of n or polynomial in the block length q^n.

Let $\mathcal{P}_d(\mathbb{K}^n)$ denote the class of degree $\leqslant d$ polynomials $f : \mathbb{F}^n \to \mathbb{F}$. Let dist denote the normalized Hamming distance. For $\mathrm{RM}_{\mathbb{K}}(n, d)$, $\eta > 0$, let

$$\ell_{\mathbb{F}}(n, d, \eta) := \max_{g:\mathbb{F}^n \to \mathbb{F}} |\{f \in \mathcal{P}_d(\mathbb{F}^n) : \mathrm{dist}(f, g) \leqslant \eta\}| .$$

Let $\mathrm{LDR}_{\mathbb{K}}(n, d)$ (short for *list decoding radius*) be the maximum ρ for which $\ell_{\mathbb{K}}(n, d, \rho - \varepsilon)$ is upper bounded by a constant depending only on $\varepsilon, |\mathbb{K}|, d$ for all $\varepsilon > 0$.

It is easy to see that $\mathrm{LDR}_{\mathbb{K}}(n, d) \leqslant \delta_{\mathbb{K}}(d)$. The difficulty lies in proving a matching lower bound. We review some previous work next. The first breakthrough result was the celebrated work of Goldreich and Levin [20] who showed that in the setting of $d = 1$ over $\mathbb{F}_2$ (Hadamard Codes) $\mathrm{LDR}_{\mathbb{F}_2}(n, 1) = \delta_{\mathbb{F}_2}(1) = 1/2$. Later, Goldreich, Rubinfield and Sudan [21] generalized the field to obtain $\mathrm{LDR}_{\mathbb{K}}(n, 1) = \delta_{\mathbb{K}}(1) = 1 - 1/|\mathbb{K}|$. In the setting of $d < |\mathbb{K}|$, Sudan, Trevisan and Vadhan [45] showed that $\mathrm{LDR}_{\mathbb{K}}(n, d) \geqslant 1 - \sqrt{2d/|\mathbb{K}|}$ improving previous work by Arora and Sudan [2], Goldreich *et al* [21] and Pellikaan and Wu [41]. Note that this falls short of the upper bound which is $\delta_{\mathbb{K}}(d)$.

In 2008, Gopalan, Klivans and Zuckerman [26] showed that $\mathrm{LDR}_{\mathbb{F}_2}(n, d) = \delta_{\mathbb{F}_2}(d)$.They posed the following conjecture.

▶ **Conjecture 1** ([26]). *For fixed d and finite field $\mathbb{K}$, $\mathrm{LDR}_{\mathbb{K}}(n, d) = \delta_{\mathbb{K}}(d)$.*

It is believed [26, 25] that the hardest case is the setting of small d. An important step in this direction was taken in [25] that considered quadratic polynomials and showed that $\mathrm{LDR}_{\mathbb{K}}(n, 2) = \delta_{\mathbb{K}}(2)$ for all fields $\mathbb{K}$ and thus proved the conjecture for $d = 2$. Recently, Bhowmick and Lovett [14] resolved the conjecture for prime $\mathbb{K}$.

Our main result for list decoding is a resolution of Conjecture 1.

▶ **Theorem 2.** *Let $\mathbb{K}$ be a finite field. Let $\varepsilon > 0$ and $d, n \in \mathbb{N}$. Then,*

$$\ell_{\mathbb{K}}(d, n, \delta_{\mathbb{K}}(d) - \varepsilon) \leqslant c_{|\mathbb{K}|,d,\varepsilon}.$$

Thus,

$$\mathrm{LDR}_{\mathbb{K}}(n, d) = \delta_{\mathbb{K}}(d).$$

▶ Remark (Algorithmic Implications). Using the blackbox reduction of algorithmic list decoding to combinatorial list decoding in [26] along with Theorem 18, for fixed finite fields, d and $\varepsilon > 0$, we now have list decoding algorithms in both the global setting (running time polynomial in $|\mathbb{K}|^n$) and the local setting (running time polynomial in n^d).

▶ Remark. Note that the bound on the list size in Theorem 2 depends on $|\mathbb{K}|$. The recent work of Bhowmick and Lovett [13] shows that this is not necessary when $d < p$. It is an open question to show this for general fields. Our dichotomy result for polynomials on general fields can't be used in their proof because it only holds for polynomials mapping to $\mathbb{K}$, not for the more general non-classical polynomials.

1.2.2 Algorithmic polynomial decomposition

Consider the following family of properties of functions over a finite field $\mathbb{K}$.

▶ **Definition 3.** Given a positive integer k, a vector of positive integers $\mathbf{\Delta} = (\Delta_1, \Delta_2, \ldots, \Delta_k)$ and a function $\Gamma : \mathbb{K}^k \to \mathbb{K}$, we say that a function $P : \mathbb{K}^n \to \mathbb{K}$ is $(k, \mathbf{\Delta}, \Gamma)$-*structured* if there exist polynomials $P_1, P_2, \ldots, P_k : \mathbb{K}^n \to \mathbb{K}$ with each $\deg(P_i) \leqslant \Delta_i$ such that for all $x \in \mathbb{K}^n$,

$$P(x) = \Gamma(P_1(x), P_2(x), \ldots, P_k(x)).$$

The polynomials $P_1, \ldots, P_k$ are said to form a $(k, \mathbf{\Delta}, \Gamma)$-*decomposition*.

For instance, an n-variate polynomial over the field $\mathbb{K}$ of total degree d factors nontrivially exactly when it is $(2, (d-1, d-1), \mathsf{prod})$-structured where $\mathsf{prod}(a, b) = a \cdot b$. We shall use the term *degree-structural property* to refer to a property from the family of $(k, \mathbf{\Delta}, \Gamma)$-structured properties.

The problem here is, for arbitrary fixed $k, \mathbb{K}, \mathbf{\Delta}, \Gamma$, given a polynomial, decide efficiently if it is degree structural and if yes, output the decomposition. An efficient algorithm for the above would imply a (deterministic) poly(n)-time algorithm for factoring an n-variate polynomial of degree d over $\mathbb{K}$. Also, it implies a polynomial time algorithm for deciding whether a d-dimensional tensor over $\mathbb{K}$ has rank at most r. Also, it would give polynomial time algorithms for a wide range of problems not known to have non-trivial solutions previously, such as whether a polynomial of degree d can be expressed as $P_1 \cdot P_2 + P_3 \cdot P_4$ where each P_1, P_2, P_3, P_4 are of degree $d-1$ or less.

This problem was solved for prime $\mathbb{K} = \mathbb{F}$, satisfying $d < p$ by Bhattacharyya [6] and later for all d and prime $\mathbb{F}$ by Bhattacharyya, Hatami and Tulsiani [12]. Our main result in this line of work establishes this for all fixed finite fields.

▶ **Theorem 4.** *For every finite field $\mathbb{K}$, positive integers k and d, every vector of positive integers $\mathbf{\Delta} = (\Delta_1, \Delta_2, \ldots, \Delta_k)$ and every function $\Gamma : \mathbb{K}^k \to \mathbb{K}$, there is a deterministic algorithm $\mathcal{A}_{\mathbb{K},d,k,\mathbf{\Delta},\Gamma}$ that takes as input a polynomial $P : \mathbb{K}^n \to \mathbb{K}$ of degree d that runs in time polynomial in n, and outputs a $(k, \mathbf{\Delta}, \Gamma)$-decomposition of P if one exists while otherwise returning* NO.

1.2.3 Testing affine-invariant properties

The goal of property testing, as initiated by [15, 4] and defined formally by [42, 22], is to devise algorithms that query their input a very small number of times while correctly deciding whether the input satisfies a given property or is "far" from satisfying it. A property is called *testable* if the query complexity can be made independent of the size of the input.

More precisely, we use the following definitions. Let $[R]$ denote the set $\{1, \ldots, R\}$. Given a property $\mathcal{P}$ of functions in $\{\mathbb{K}^n \to [R] \mid n \in \mathbb{Z}_{\geqslant 0}\}$, we say that $f : \mathbb{K}^n \to [R]$ is ε-*far* from $\mathcal{P}$ if

$$\min_{g \in \mathcal{P}} \mathbf{Pr}_{x \in \mathbb{K}^n}[f(x) \neq g(x)] > \varepsilon,$$

and we say that it is ε-*close* otherwise.

▶ **Definition 5** (Testability)**.** A property $\mathcal{P}$ is said to be *testable* (with one-sided error) if there are functions $q : (0,1) \to \mathbb{Z}_{>0}$, $\delta : (0,1) \to (0,1)$, and an algorithm T that, given as input a parameter $\varepsilon > 0$ and oracle access to a function $f : \mathbb{K}^n \to [R]$, makes at most $q(\varepsilon)$

queries to the oracle for f, always accepts if $f \in \mathcal{P}$ and rejects with probability at least $\delta(\varepsilon)$ if f is ε-far from $\mathcal{P}$. If, furthermore, q is a constant function, then $\mathcal{P}$ is said to be *proximity-obliviously testable (PO testable).*

The term proximity-oblivious testing is coined by Goldreich and Ron in [24]. As an example of a testable (in fact, PO testable) property, let us recall the famous result by Blum, Luby and Rubinfeld [15] which initiated this line of research. They showed that linearity of a function $f : \mathbb{K}^n \to \mathbb{K}$ is testable by a test which makes 3 queries. This test accepts if f is linear and rejects with probability $\Omega(\varepsilon)$ if f is ε-far from linear.

Linearity, in addition to being testable, is also an example of a *linear-invariant* property. We say that a property $\mathcal{P} \subseteq \{\mathbb{K}^n \to [R]\}$ is linear-invariant if it is the case that for any $f \in \mathcal{P}$ and for any $\mathbb{K}$-linear transformation $L : \mathbb{K}^n \to \mathbb{K}^n$, it holds that $f \circ L \in \mathcal{P}$. Similarly, an *affine-invariant* property is closed under composition with affine transformations $A : \mathbb{K}^n \to \mathbb{K}^n$ (an affine transformation A is of the form $L + c$ where L is $\mathbb{K}$-linear and $c \in \mathbb{K}$). The property of a function $f : \mathbb{K}^n \to \mathbb{K}$ being affine is testable by a simple reduction to [15], and is itself affine-invariant. Other well-studied examples of affine-invariant (and hence, linear-invariant) properties include Reed-Muller codes [4, 3, 19, 42, 1] and Fourier sparsity [27]. In fact, affine invariance seems to be a common feature of most interesting properties that one would classify as "algebraic". Kaufman and Sudan in [39] made explicit note of this phenomenon and initiated a general study of the testability of affine-invariant properties (see also [23]).

Our main theorem for testing is a very general positive result:

▶ **Theorem 6** (Main testing result). *Let $\mathcal{P} \subseteq \{\mathbb{K}^n \to [R]\}$ be an affine-invariant property that is t, w-lightly locally characterized, where t, R, w, and* $\mathrm{char}(\mathbb{K})$ *are fixed positive integers. Then, $\mathcal{P}$ is PO testable with t queries.*

We are yet to define several terms in the above claim, but as we will see, the weight restriction is trivial when the field size is bounded. This yields the following characterization.

▶ **Theorem 7** (Testing result for fixed fields). *Let $\mathcal{P} \subseteq \{\mathbb{K}^n \to [R]\}$ be an affine-invariant property, where $R \in \mathbb{Z}^+$ and field $\mathbb{K}$ are fixed. Then, $\mathcal{P}$ is PO testable with t queries if and only if $\mathcal{P}$ is t-locally characterized.*

Previously, [9] (building on [8, 11, 10]) proved Theorem 6 in the case that $\mathbb{K}$ is of fixed prime order using higher-order Fourier analytic techniques. We note that other recent results on 2-sided testability of affine-invariant properties over fixed prime-order fields [35, 51] can also be similarly extended to non-prime fields but we omit their description here.

1.2.3.1 Local Characterizations

For a PO testable property $\mathcal{P} \subset \{\mathbb{K}^n \to [R]\}$ of query complexity t, if a function $f : \mathbb{K}^n \to [R]$ does not satisfy $\mathcal{P}$, then by Definition 5, the tester rejects f with positive probability. Since the test always accepts functions with the property, there must be t points $a_1, \ldots, a_t \in \mathbb{K}^n$ that form a witness for non-membership in $\mathcal{P}$. These are the queries that cause the tester to reject. Thus, denoting $\sigma = (f(a_1), \ldots, f(a_t)) \in [R]^t$, we say that $\mathcal{C} = (a_1, a_2, \ldots, a_t; \sigma)$ forms a *t-local constraint* for $\mathcal{P}$. This means that whenever the constraint is violated by a function g, i.e., $(g(a_1), \ldots, g(a_t)) = \sigma$, we know that g is not in $\mathcal{P}$. A property $\mathcal{P}$ is *t-locally characterized* if there exists a collection of t-local constraints $\mathcal{C}_1, \ldots, \mathcal{C}_m$ such that $g \in \mathcal{P}$ if and only if none of the constraints $\mathcal{C}_1, \ldots, \mathcal{C}_m$ are violated. It follows from the above discussion that if $\mathcal{P}$ is PO testable with q queries, then $\mathcal{P}$ is t-locally characterized.

For an affine-invariant property, constraints can be defined in terms of affine forms, since the affine orbit of a constraint is also a constraint. So, we can describe each t-local constraint

$\mathcal{C}$ as $(A_1, \ldots, A_t; \sigma)$, where for every $i \in [t]$, $A_i(X_1, \ldots, X_t) = X_1 + \sum_{j=2}^{t} c_{i,j} X_j$ for some $c_{i,j} \in \mathbb{K}$ is an affine form over $\mathbb{K}$. We define the *weight* wt of an element $c \in \mathbb{K}$ as $\sum_{k=1}^{r} |c_k|$, where c is viewed as an r-dimensional vector $(c_1, \ldots, c_r)$ with each c_i in the base prime field[2] $\mathbb{F}$ with respect to a fixed arbitrary basis. The *weight of an affine form* A_i to be $\sum_{j=2}^{m} \mathsf{wt}(c_{i,j})$ for $c_{i,j}$ as above. A constraint is said to be of weight w if all its affine forms are of weight at most w, and a property $\mathcal{P}$ is said to be t, w-lightly locally characterized if there exist t-local constraints $\mathcal{C}_1, \ldots, \mathcal{C}_m$, each of weight at most w that characterize $\mathcal{P}$.

Theorem 6 asserts that if $\mathcal{P}$ has a light local characterization, then it is testable. There can exist many local characterizations of a property, and for the theorem to apply, it is only necessary that one such characterization be of bounded weight. Moreover, we can choose the basis with which to describe $\mathbb{K}$ over $\mathbb{F}$. On the other hand, some restriction in addition to local characterization is needed, as Ben-Sasson et al. [5] show that there exist affine-invariant locally characterized properties of functions $f : \mathbb{F}_{2^n} \to \mathbb{F}_2$ that require super-constant query complexity to test.

Another interesting observation is that if a property has a local characterization of bounded weight, then it has a local *single orbit characterization*, in the language of [39]. For linear[3] affine-invariant properties, [39] shows that any local single orbit characterized property is testable. Hence, our result is weaker than [39] in this aspect, though our Theorem 6 allows non-linear properties. It is an interesting open question as to whether dual-BCH codes and, more generally, sparse affine-invariant codes that were shown to be locally single orbit characterized in [36] and [31] respectively also have local characterizations of bounded weight. It is also an open problem to describe a testable property $\mathcal{P} \subseteq \{\mathbb{F}_{2^n} \to \mathbb{F}_2\}$ that does not have a local characterization of bounded weight.

1.3 Parallel Work

Subsequent to our first public version of this work [7], Bhowmick and Lovett [13] proved Theorem 2 for all finite fields, even when the field size is growing with n, but assuming that the field characteristic is larger than the order of the Reed-Muller code. They focus more on handling growing field size instead of arbitrary field characteristic, so their techniques are substantially different.

1.4 Our Techniques

1.4.1 New Ingredients

Our starting point is the observation that $\mathbb{K}$ is an r-dimensional vector space over $\mathbb{F}$. Thus, we can view a function $Q : \mathbb{K}^n \to \mathbb{K}$ as determined by a collection of functions $P_1, \ldots, P_r : \mathbb{K}^n \to \mathbb{F}$ where $\mathbb{K}^n$ is viewed as $\mathbb{F}^{rn}$. However, it is incorrect to suppose $P_1, \ldots, P_r$ are independent as they are generated by the same polynomial over $\mathbb{K}$.

Indeed, in our first dichotomy theorem, we want to deduce structural information about P just from the fact that P_1 is biased. Although we can't directly prove that biased P_1 implies biased P_i for all $i \in [r]$, we show that biased DP_1 implies biased DP_i for all $i \in [r]$, where for a polynomial Q of degree d, $DQ(h_1, \ldots, h_d)$ is the iterated derivative of Q in directions $h_1, \ldots, h_d$. Because d is the total degree of the polynomial, the iterated derivative

[2] If $x \in \mathbb{F}$, $|x|$ is the obvious element of $\{0, 1, \ldots, |\mathbb{F}| - 1\}$.
[3] These are properties of functions $f : \mathbb{K}^n \to \mathbb{F}$, where $\mathbb{F}$ is a subfield of $\mathbb{K}$, for which $f, g \in \mathcal{P}$ implies $\alpha f + \beta g \in \mathcal{P}$ for any $\alpha, \beta \in \mathbb{F}$.

is multilinear in $h_1, \ldots, h_d$. The multilinearity allows us to relate the structure of DP_1 to the rest of the DP_is. The same strategy was used in [13]. We then follow the steps of [48] to integrate each of the DP_i's. We show that all of them are functions of the same collection of *non-classical polynomials*. Here, a non-classical polynomial $Q(x_1, \ldots, x_n)$ is determined by a set of monomials $ax_1^{i_1}x_2^{i_2}\cdots x_n^{i_n}$ where a and the multiplication is in $\mathbb{K}$; the evaluation of each monomial is then mapped to $(\mathbb{Z}/p^{k+1}\mathbb{Z})^r$ for some integer $k \geqslant 0$ and the final output is an integer linear combination of them. Note that unlike our definition, in the definition of non-classical polynomials over $\mathbb{F}^n$ by [48], the multiplicative structure of the field is never used. Also, our non-classical polynomials are a strict subset of the functions $P : \mathbb{K}^n \to \mathbb{T}^r$ which identically vanish after $d+1$ derivatives. In fact, if we had used the latter notion as defining non-classical polynomials, our theorem would have been quite straightforward.

Over constant-sized fields $\mathbb{K}$, it is more economical to write out P as determined by $P_1, \ldots, P_r$ and treat them as independent. Then, we have reduced the problem to studying polynomials mapping to $\mathbb{F}$. However, even in this setting, we cannot totally ignore the multiplicative structure of $\mathbb{K}$. To see why, recall the question of testing affine-invariant properties. When $\mathbb{K}$ is of bounded order, we can view any one-sided test as examining the restriction of the input function on a random K-dimensional affine subspace of $\mathbb{K}^n$, for some constant integer K. In other words, the test will evaluate the input function at elements of the set $H = \{x + \sum_{i=1}^{K} a_i y_i : a_1, \ldots, a_K \in \mathbb{K}\}$ for some $x, y_1, \ldots, y_K \in \mathbb{K}$. Clearly, H is not an affine subspace of $\mathbb{F}^{rn}$ because of $\mathbb{K}$'s multiplicative structure. An important component of the higher-order Fourier analytic approach is to show that any "sufficiently pseudorandom" collection of polynomials is equidistributed on H, and the proof of this fact in [9] crucially uses that H is a subspace of a vector space over a prime field. In our work, we show a strong equidistribution theorem (Theorem 36) that holds when H is an affine subspace of $\mathbb{K}^n$.

A different place where the multiplicative structure of $\mathbb{K}$ rears its head is a key *Degree Preserving Lemma* of [9]. Informally, if $P_1, \ldots, P_C$ form a "sufficiently pseudorandom" collection of polynomials and $F(x) = \Gamma(P_1(x), \ldots, P_C(x))$ is a polynomial of degree d where Γ is an arbitrary composition function, then for any other collection of polynomials $Q_1, \ldots, Q_C$ where $\deg(Q_i) \leqslant \deg(P_i)$ for every i, $G(x) = \Gamma(Q_1(x), \ldots, Q_C(x))$ also has degree $\leqslant d$. The lemma is crucially used for the analysis of the Reed-Muller list decoding bound in [14] and the polynomial decomposition algorithm in [6, 12]. Its proof goes via showing that if all $(d+1)$ iterated derivatives of $F : \mathbb{K}^n \to \mathbb{K}$ vanish, then so must all $(d+1)$ iterated derivatives of $G : \mathbb{K}^n \to \mathbb{K}$. However, for $\mathbb{K}$ that is of size p^2 or more, this only implies a bound on the weight degree of G, not on its degree.

We resolve this issue by giving a different and more transparent proof of the Degree Preserving Lemma, which actually holds in a much more general setting. Using the above notation, we prove that if $F : \mathbb{K}^n \to \mathbb{K}$ satisfies some locally characterized property $\mathcal{P}$, then $G : \mathbb{K}^n \to \mathbb{K}$ does also. Since due to a work of Kaufman and Ron [38], we know that degree is locally characterized, our desired result follows. Our new proof uses our strong equidistribution theorem on affine subspaces of $\mathbb{K}^n$. An interesting point to note is that both the equidistribution theorem and the degree preserving lemma work only assuming that the field characteristic is constant and that the involved affine constraints are of bounded weight, without any assumption on the field size.

1.4.2 Reed-Muller codes

For a received word $g : \mathbb{K}^n \to \mathbb{K}$ our goal is to upper bound $|\{f \in \mathcal{P}_d : \mathrm{dist}(f, g) \leqslant \eta\}|$, where $\eta = \delta_{\mathbb{K}}(d) - \varepsilon$ for some $\eta > 0$ and $\mathcal{P}_d$ is the class $\{Q : \mathbb{K}^n \to \mathbb{K} : \deg(Q) \leqslant d\}$. The proof technique is similar in structure as [14]. We apply the weak regularity lemma (Corollary 40)

to the received word $g : \mathbb{K}^n \to \mathbb{K}$ and reduce the problem to a structured word $g' : \mathbb{K}^n \to \mathbb{K}$. More specifically, whenever $\mathrm{dist}(f, g) \leqslant \eta$, we have $\mathrm{dist}(f, g') \leqslant \eta + \varepsilon/2$. From here, we first express each function $f : \mathbb{K}^n \to \mathbb{K}$ as a linear combination of functions $f' : \mathbb{K}^n \to \mathbb{F}$. It can be then shown that the analysis in [14] works for functions $f' : \mathbb{K}^n \to \mathbb{F}$. A naive recombination of the $f' : \mathbb{K}^n \to \mathbb{F}$ to $f : \mathbb{K}^n \to \mathbb{K}$ gives us useful bounds only when $d < \mathrm{char}(|\mathbb{F}|)$. To circumvent this problem, we use our improved degree preserving theorem. This is crucial to our analysis as the technique of [14] can be used only to analyze the weight degree of polynomials which is not enough for the argument to work for arbitrary d and $|\mathbb{K}|$.

1.4.3 Polynomial decomposition

The algorithm and its analysis follows the lines of [6, 12]. Given a polynomial $P : \mathbb{K}^n \to \mathbb{K}$ (where $|\mathbb{K}|$ is bounded), we consider the collection of polynomials $\{\mathrm{Tr}(\alpha_1 P), \ldots, \mathrm{Tr}(\alpha_r P)\}$ where $\alpha_1, \ldots, \alpha_r \in \mathbb{K}$ are linearly independent. We regularize this collection into a pseudo-random polynomial factor and set one variable to 0 such that the degrees of the polynomials do not change. We then recursively solve the problem on $n - 1$ variables and then apply a lifting procedure to get a decomposition for the original problem. A naive analysis of the lifting procedure over non-prime fields requires that $\deg(P) < \mathrm{char}(\mathbb{F})$. In order to get around this, we use our improved degree preserving theorem which applies for arbitrary degrees.

1.4.4 Testing affine-invariant properties

Suppose $\mathcal{P} \subseteq \{\mathbb{K}^n \to [R]\}$ is a locally characterized affine-invariant property (where R and $\mathrm{char}(\mathbb{K})$ are bounded but $n|\mathbb{K}|$ is growing). Our proof follows the lines of [11, 10, 9]. Suppose f is far from $\mathcal{P}$. We first identify a low-rank function close to f in an appropriate Gowers norm which also contains the violation that f contains. Here, low rank is with respect to a collection $\mathcal{B}$ of non-classical polynomials mapping to $\mathbb{T}$. We then investigate the distribution of $\mathcal{B}$ on the affine constraint that f violates. Since these are affine with respect to $\mathbb{K}^n$, we need to use our strong equidistribution theorem. The rest of the proof proceeds along the same lines as [9].

Because the proof of Theorem 6 is very analogous to that in [9] (except for the use of the new equidistribution theorem) and requires significant additional notation, we omit it here.

1.5 Some Open Questions

We conclude the introduction by giving a list of open questions suggested by this line of work:

- In our dichotomy result, can we show that if $P : \mathbb{K}^n \to \mathbb{K}$ is a polynomial of *weight degree* d, then either a character of it is unbiased or P is expressible in terms of other polynomials of weight degrees less than d? In the current form of the theorem, d is the degree of P.
- Can we show the dichotomy theorem for non-classical polynomials P? Together with the first item, we would then be able to iteratively regularize collections of polynomials over finite fields of growing size. Such a procedure would help resolve the list decoding radius for Reed-Muller codes over such fields, for instance.
- Can we improve the bound on the list size for Reed-Muller codes, compared to what we obtain in Theorem 2 or what is obtained in [13]?
- Can we use higher-order Fourier analysis to investigate the list-decoding radius for other codes, most notably, the Reed-Solomon code or the lifted Reed-Muller codes [32]?

- Is the notion of bounded weight characterization of affine-invariant properties an artifact of our proof or is it indeed linked to testability?

1.6 Organization

We formally define some notions like polynomials, bias, Gowers norm, and rank in Section 2. In Section 3, we show the bias versus rank dichotomy for non-classical polynomials. In Section 4, we show the equidistribution results for polynomials over subspaces. The next two sections, Sections 5 and 6, describe how the new tools can be used to prove the results for Reed-Muller list-decoding radius and algorithmic polynomial decomposition.

2 Preliminaries

Fix a prime number $p \geqslant 2$. Let $\mathbb{F}$ be the finite field of order p, and let $\mathbb{K}$ be a finite field of characteristic p. Let r denote the dimension of $\mathbb{K}$ as a vector space over $\mathbb{F}$; so, $|\mathbb{K}| = p^r$. Note that r is a parameter and may not be held constant.

Let $\mathrm{Tr} : \mathbb{K} \to \mathbb{F}$ denote the trace function: $\mathrm{Tr}(x) = x + x^p + x^{p^2} + \cdots + x^{p^{r-1}}$ for $x \in \mathbb{K}$. Recall that $\{x \to \mathrm{Tr}(ax) : a \in \mathbb{K}\}$ is in bijection with the set of all $\mathbb{F}$-linear maps from $\mathbb{K}$ to $\mathbb{F}$.

For every $\mathbb{K}$ of order r over $\mathbb{F}$, fix a choice of r linearly independent field elements $\alpha_1, \alpha_2, \ldots, \alpha_r \in \mathbb{K}$. Then, there exists a dual basis, $\beta_1, \beta_2, \ldots, \beta_r \in \mathbb{K}$ such that any $x \in \mathbb{K}$ can be written as

$$x = \sum_{i=1}^{r} \beta_i \mathrm{Tr}(\alpha_i x) \tag{1}$$

In particular, $\mathrm{Tr}(\alpha_i \beta_j)$ equals 1 if $i = j$ and 0 otherwise.

Let $\mathbb{T}$ denote the torus $\mathbb{R}/\mathbb{Z}$. This is an abelian group under addition. . For an integer $k \geqslant 0$, let $\mathbb{U}_k := \frac{1}{p^k}\mathbb{Z}/\mathbb{Z}$. Note that $\mathbb{U}_k$ is a subgroup of $\mathbb{T}$. Let $\iota : \mathbb{F} \to \mathbb{U}_1$ be the bijection $\iota(a) = \frac{|a|}{p} \pmod 1$. This map naturally extends to $\kappa : \mathbb{K} \to \mathbb{U}_1[Z_1, Z_2, \ldots Z_r]$ where $Z_1, \ldots, Z_r$ are formal variables, by the bijection $\kappa(x) = \sum_{i=1}^{r} \iota(\mathrm{Tr}(\alpha_i x)) \cdot Z_i$.

Let $e : \mathbb{T} \to \mathbb{C}$ be the function $e(x) = e^{2\pi i x}$. By abuse of notation, we sometimes write $e(x)$ for $x \in \mathbb{F}$ to mean $e(\iota(x))$.

2.1 Classical and Non-Classical Polynomials

We start by defining *classical polynomials*.

▶ **Definition 8.** (Classical polynomials over $\mathbb{K}$). We say that $P : \mathbb{K}^n \to \mathbb{K}$ is a *classical polynomial* of *degree* $\leqslant d$ if there exist coefficients $\{c_{i_1,\ldots,i_n} \in \mathbb{K} : i_1, \ldots, i_n \geqslant 0\}$ such that for all $x \in \mathbb{K}^n$:

$$P(x) = \sum_{i_1,\ldots,i_n \geqslant 0, \sum_j i_j \leqslant d} c_{i_1,\ldots,i_n} x_1^{i_1} \cdots x_n^{i_n}$$

where $c_{i_1,\ldots,i_n}, x_1, \ldots, x_n \in \mathbb{K}$.

As discussed above, there is a bijection $\kappa : \mathbb{K} \to \mathbb{T}[Z_1, \ldots, Z_r]$ where $Z_1, \ldots, Z_r$ are formal variables. This bijection carries a classical polynomial $P : \mathbb{K}^n \to \mathbb{K}$ of degree $\leqslant d$ to a

function of the form:

$$\kappa(P(x)) = \sum_{j=1}^{r} \sum_{\substack{0 \leqslant i_1,\ldots,i_n: \\ \sum_j i_j \leqslant d}} \left(\frac{|\mathrm{Tr}(\alpha_j \cdot c_{i_1,\ldots,i_n} \cdot x_1^{i_1} \cdots x_n^{i_n})|}{p} \pmod 1 \right) \cdot Z_j$$

where $c_{i_1,\ldots,i_n} \in \mathbb{K}$. *Non-classical polynomials* are defined to map to $\mathbb{T}[Z_1, \ldots, Z_r]$ and have a similar representation.

▶ **Definition 9** (Non-classical polynomials over $\mathbb{K}$)**.** We say that $Q : \mathbb{K}^n \to \mathbb{T}[Z_1, \ldots, Z_r]$ is a *non-classical polynomial* of *degree* $\leqslant d$ and *height* $\leqslant k$ if Q can be written as:

$$Q(x) = \sum_{j=1}^{r} \left(\gamma_j + \sum_{\ell=0}^{k} \sum_{\substack{0 \leqslant i_1,\ldots,i_n < p^r: \\ \sum_j i_j \leqslant d-\ell(p-1)}} \left(\frac{|\mathrm{Tr}(\alpha_j \cdot c_{i_1,\ldots,i_n,\ell} \cdot x_1^{i_1} \cdots x_n^{i_n})|}{p^{\ell+1}} \pmod 1 \right) \right) \cdot Z_j$$

for some $c_{i_1,\ldots,i_n,\ell} \in \mathbb{K}$ and $\gamma_j \in \mathbb{T}$.

Crucially, note that the coefficients $c_{i_1,\ldots,i_n,\ell}$ do not depend on j. Also, observe that non-classical polynomials of height 0 correspond to classical polynomials and that if $\mathbb{K} = \mathbb{F}$, this definition is identical to the one in [48].

In many parts of this paper, we will speak of non-classical polynomials $P : \mathbb{K}^n \to \mathbb{T}$. More precisely, this means that we identify $\mathbb{K}$ with $\mathbb{F}^{rn}$, and P is actually a non-classical polynomial over $\mathbb{F}$. In particular, it has the form:

$$P(x_1, \ldots, x_n) = \alpha + \sum_{\ell=0}^{k} \sum_{\substack{0 \leqslant d_{i,j} < p \ \forall i \in [n], j \in [r]: \\ \sum_{i=1}^{n} \sum_{j=1}^{r} d_{i,j} \leqslant d-k(p-1)}} \frac{c_{d_{1,1},\ldots,d_{n,r},k} \prod_{i=1}^{n} \prod_{j=1}^{r} |\mathrm{Tr}(\alpha_j x_i)|^{d_{i,j}}}{p^{\ell+1}} \pmod 1$$

where $\alpha \in \mathbb{T}$ and $c_{d_{1,1},\ldots,d_{n,r}} \in \{0, 1, \ldots, p-1\}$

A particular type of classical polynomial plays an important role in our analysis.

▶ **Definition 10** (Classical, symmetric, multilinear (CSM) forms)**.** We say that a polynomial $T : (\mathbb{K}^n)^d \to \mathbb{K}$ is in $\mathrm{CSM}_d(\mathbb{K}^n)$ if $T(h_1, \ldots, h_d)$ is of the form

$$T(h_1, \ldots, h_d) = \sum_{i_1,\ldots,i_d \in [n]} c_{\{i_1,\ldots,i_d\}} h_{1,i_1} \cdots h_{d,i_d}$$

where $c_{\{i_1,\ldots,i_d\}} \in \mathbb{K}$ and $h_1, \ldots, h_d \in \mathbb{K}^n$. T satisfies the following properties

1. Multilinear: Each term in T has the form above.
2. Symmetric: T is invariant with regards to permutations of $h_1, \ldots, h_d$.
3. Classical: $T(h_1, \ldots, h_d)$ vanishes whenever at least p of the $h_1, \ldots, h_d \in \mathbb{K}^n$ are equal.

2.2 Additive Derivatives and Weight Degree

Although polynomials are defined in the above section in terms of their global representation, we will often be interested in local constraints obeyed by functions.

▶ **Definition 11** (Additive derivative and Weight degree)**.** Given a function $f : \mathbb{K}^n \to \mathbb{T}[Z_1, \ldots, Z_r]$, its *additive derivative in direction* $h \in \mathbb{K}^n$ is $D_h f : \mathbb{K}^n \to \mathbb{T}[Z_1, \ldots, Z_r]$, given by

$$D_h f(x) = f(x + h) - f(x).$$

A function $P : \mathbb{K}^n \to \mathbb{T}[Z_1, \ldots, Z_r]$ is said to have *weight degree* $\leqslant w$ if for all $x, h_1, h_2, \ldots, h_{w+1} \in \mathbb{K}^n$,

$$D_{h_1} D_{h_2} \cdots D_{h_{w+1}} P(x) = 0. \tag{2}$$

If $f : \mathbb{K}^n \to \mathbb{T}[Z_1, \ldots, Z_r]$ and $f(x) = \sum_{i=1}^{r} f_i(x) \cdot Z_i$, then it is clear that f has weight degree $\leqslant w$ if and only if each $f_i : \mathbb{K}^n \to \mathbb{T}$ is a non-classical polynomial of degree $\leqslant w$ (in the sense of [48]). This is because for functions mapping from a vector space over $\mathbb{F}$ to $\mathbb{T}$, the notion of degree and weight degree coincide. In particular:

▶ **Fact 12.** *The degree and weight degree of any non-classical polynomial $P : \mathbb{K}^n \to \mathbb{T}$ are equal.*

But what about the relation between degree and weight degree for functions mapping to $\mathbb{T}[Z_1, \ldots, Z_r]$? Here, we can make two remarks.

▶ Remark. As mentioned above, a bound of w for the weight degree of a function $f = \sum_i f_i \cdot Z_i$ means that each f_i is individually of (weight) degree $\leqslant w$, but it does not impose any relationship whatsoever between the different f_i's. On the other hand, in Definition 9, the different f_i's are all determined by the same set of coefficients in $\mathbb{K}$.

▶ Remark. If $P : \mathbb{K}^n \to \mathbb{T}[Z_1, \ldots, Z_r]$ is a non-classical polynomial, then its weight degree is the maximum weight degree of any of its terms, where the weight degree of a term $\frac{\mathrm{Tr}(\alpha_j c_{i_1,\ldots,i_n} \ell x_1^{i_1} \cdots x_n^{i_n})}{p^{\ell+1}}$ is $\ell(p-1) + \mathsf{wt}(i_1) + \mathsf{wt}(i_2) + \cdots + \mathsf{wt}(i_n)$ and $\mathsf{wt}(i)$ for $0 \leqslant i < p^r$ is the sum of the r digits of i in its p-ary expansion. Hence, if every individual degree i_k is less than p, then the weight degree equals the degree of the polynomial. Also, clearly, the weight degree is always at most the degree for any non-classical polynomial.

2.3 Bias and Gowers norm

▶ **Definition 13.** The *bias* of a function $f : \mathbb{K}^n \to \mathbb{K}$ is defined as $\mathrm{bias}(P) = |\mathbb{E}_{x \in \mathbb{K}^n} e(\mathrm{Tr}(P(x)))|$.

Here, we could have used any non-trivial additive character instead of the trace, but we choose this definition for concreteness (without any loss of generality). The *Gowers norm* of a function measures the bias of its iterated derivative.

▶ **Definition 14** (Gowers norm)**.** Given a function $P : \mathbb{K}^n \to \mathbb{K}$ and an integer $d \geqslant 1$, the *Gowers norm of order d* for P is given by

$$\|P\|_{U^d} = \left| \mathop{\mathbb{E}}_{h_1,\ldots,h_d,x \in \mathbb{K}^n} [e(D_{h_1} \cdots D_{h_d} \mathrm{Tr}(P(x)))] \right|^{1/2^d}.$$

Note that as $\|f\|_{U^1} = \mathrm{bias}(f)$ the Gowers norm of order 1 is only a semi-norm. However for $d > 1$, it is not difficult to show that $\|\cdot\|_{U^d}$ is indeed a norm. Also, note that $\|P\|_{U^d} \geqslant \mathrm{bias}(P)$ for any $d \geqslant 1$.

We can also write the Gowers norm in a slightly different way which will be convenient for us.

▶ **Definition 15** (Derivative polynomial)**.** If $P : \mathbb{K}^n \to \mathbb{K}$ is a classical polynomial of degree d (i.e., some term of P has total degree exactly d), then let the *derivative polynomial* be $DP : \mathbb{K}^{nd} \to \mathbb{K}$ defined as $DP(h_1, \ldots, h_d) = D_{h_1} D_{h_2} \cdots D_{h_d} P(x)$, which is independent of x.

▶ **Lemma 16.** *Let $P : \mathbb{K}^n \to \mathbb{K}$ be a classical polynomial of degree d. Then, $DP \in \mathrm{CSM}_d(\mathbb{K}^n)$, and $\|P\|_{U^d}^{2^d} = \mathrm{bias}(DP)$.*

Note that if d were the weight degree instead of the degree in Lemma 16, then DP would be multilinear in the sense that for any $i \in [d]$, $DP(h_i + h'_i, (h_j)_{j\neq i}) = DP(h_i, (h_j)_{j\neq i}) + DP(h'_i, (h_j)_{j\neq i})$, but individual variables could have degree more than 1 (any power of p) in DP and so could not be in $\mathrm{CSM}_d(\mathbb{K}^n)$ according to Definition 10.

2.4 Rank

▶ **Definition 17.** Let $P : \mathbb{K}^n \to \mathbb{K}$ be a classical polynomial of weight degree w. The $\mathbb{K}$-*rank* of P is the smallest integer c such that there exist functions $Q_1, \ldots, Q_c : \mathbb{K}^n \to \mathbb{T}[Z_1, \ldots, Z_r]$ of weight degree $< w$ and a function $\Gamma : \mathbb{T}[Z_1, \ldots, Z_r]^c \to \mathbb{K}$ such that $P(x) = \Gamma(Q_1(x), \ldots, Q_c(x))$.

We will often restrict to the case when p and r are constant. In this case, we can look at the collection of polynomials $\mathcal{P} = \{\mathrm{Tr}(\alpha_1 P), \ldots, \mathrm{Tr}(\alpha_r P)\}$. These are non-classical polynomials of degree w from $\mathbb{F}^{rn}$ to $\mathbb{T}$. Then, upto a factor of r, the minimum $\mathbb{F}$-rank of any nonzero linear combination of the polynomials in $\mathcal{P}$ is at most the $\mathbb{K}$-rank of P.

3 Inverse Theorem for Classical Polynomials

In this section, we establish the Gowers Inverse theorem for polynomial phases over growing field sizes.

▶ **Theorem 18.** *Let $d, p, r, s \in \mathbb{N}$, and $\mathbb{K}$ be a field extension of $\mathbb{F} = \mathbb{F}_p$ with $[\mathbb{K} : \mathbb{F}] = r$. Then, there exists $c = c^{18}(d, s)$ such that the following is true. Consider any classical polynomial $P : \mathbb{K}^n \to \mathbb{K}$ of degree $\leqslant d$ such that $\|P\|_{U^d} \geqslant |\mathbb{K}|^{-s}$. Then, there exist non-classical polynomials $R_1, \ldots, R_c : \mathbb{K}^n \to \mathbb{T}[Z_1, \ldots, Z_r]$ of weight degrees $\leqslant d - 1$ and a function $\Gamma : \mathbb{T}[Z_1, \ldots, Z_r]^c \to \mathbb{K}$ such that $P = \Gamma(R_1, \ldots, R_c)$.*

For context, recall Deligne's multivariate generalization of Weil's bound which implies that if $d = \deg(P)$ is a constant, and $\mathrm{bias}(P) > |\mathbb{K}|^{-1/2}$, then P must have weight degree 0. Our theorem shows the structure of constant-degree polynomials when their bias is smaller but still lower bounded by an inverse polynomial in $\mathbb{K}$.

An immediate corollary of Theorem 18 is that:

▶ **Corollary 19.** *Let $d, p, r, s \in \mathbb{N}$, and $\mathbb{K}$ be a field extension of $\mathbb{F} = \mathbb{F}_p$ with $[\mathbb{K} : \mathbb{F}] = r$. Then, there exists $c = c^{18}(d, s)$ such that the following is true. Consider any classical polynomial $P : \mathbb{K}^n \to \mathbb{K}$ of degree and weight degree d such that $\mathrm{bias}(P) \geqslant |\mathbb{K}|^{-s}$. Then, $\mathrm{rank}(P) \leqslant c^{18}(d, s)$.*

It is open how to remove the restriction that the degree and weight degree of P are equal. We now go on to the proof of Theorem 18.

Proof. Let DP denote the derivative polynomial of P. By Lemma 16,

$$\mathrm{bias}(\mathrm{Tr}(DP)) \geqslant |\mathbb{K}|^{-s2^d}$$

If the weight degree of P is strictly less than d, we are already done. If not, then DP is a nonzero polynomial. We will show that there exist non-classical polynomials $Q_1, \ldots, Q_c : \mathbb{K}^n \to \mathbb{T}[Z_1, \ldots, Z_r]$ of weight degree $< d$ such that for every $i \in [r]$,

$$\mathrm{Tr}(\alpha_i DP) = D\Gamma_i(Q_1, \ldots, Q_c) \tag{3}$$

for some function Γ_i mapping to $\mathbb{F}$. Therefore, because trace and derivative commute and using (1):

$$DP = D\left(\sum_{i=1}^{r} \Gamma_i(Q_1, \ldots, Q_c) \cdot \beta_i\right)$$

In other words, P and $\sum_{i=1}^{r} \Gamma_i(Q_1, \ldots, Q_c) \cdot \beta_i$ differ by a polynomial of weight degree $\leqslant d-1$, proving the theorem.

Our proof for (3) will heavily use the structure of CSM forms. To this end, let us make a couple of definitions and observations about operations on CSM forms.

▶ **Definition 20** (Concatenation)**.** Let $P \in \mathrm{CSM}_k(\mathbb{K}^n)$ and $Q \in \mathrm{CSM}_\ell(\mathbb{K}^n)$ for integers $k, \ell \geqslant 1$. Then the *concatenation operator* $P * Q \in \mathrm{CSM}_{k+\ell}(\mathbb{K}^n)$ is defined as

$$P * Q(y_1, \ldots, y_{k+\ell}) = \sum_{A \subseteq [k+\ell], |A|=k} P((y_i)_{i \in A}) Q((y_i)_{i \in [k+\ell] \setminus A})$$

▶ **Lemma 21.** *Given two classical polynomials $P : \mathbb{K}^n \to \mathbb{K}$ and $Q : \mathbb{K}^n \to \mathbb{K}$, $D(P \cdot Q) = DP * DQ$.*

▶ **Definition 22** (Symmetric Power)**.** Let $d \geqslant 2$ and $P \in \mathrm{CSM}_d(\mathbb{K}^n)$, then for $m \geqslant 1$, the *symmetric power* $\mathrm{Sym}^m(P) \in \mathrm{CSM}_{md}(\mathbb{K}^n)$ is defined as

$$\mathrm{Sym}^m(P)(h_1, \ldots, h_{md}) = \sum_{\mathcal{A}} \prod_{A \in \mathcal{A}} P((h_i)_{i \in A})$$

where the sum is over all possible partitions $\mathcal{A}$ of $\{1, \ldots, md\}$ into m-equal sized subsets.

▶ Remark. Note that $d \geqslant 2$ in the definition of symmetric power. If $d = 1$, then the symmetric power need not satisfy the third condition in Definition 10 and hence may not be CSM.

▶ Remark. Below, we'll apply the symmetric power operation to the trace of a $\mathrm{CSM}_d(\mathbb{K}^n)$ form, rather than to the form directly. However, note that the trace of a CSM is also classical, symmetric and multilinear, though now mapping to $\mathbb{F}$.

Now, we continue with the main thread of our proof. Our first step shows the structure of high-bias CSM forms.

▶ **Theorem 23** (Analog of Theorem 6.6 in [48])**.** *Suppose $d \geqslant 2$ and $s \geqslant 1$. Let $T \in \mathrm{CSM}_d(\mathbb{K}^n)$ such that $\mathrm{bias}(T(h_1, \ldots, h_d)) > |\mathbb{K}|^{-s}$. Then, there exists a subspace $V \subseteq \mathbb{K}^n$ of codimension $\leqslant r^{23}(d, s)$ such that restricted to V^d, $\mathrm{Tr}(T)$ is a linear combination over $\mathbb{F}$ of at most $t^{23}(d, s)$ expressions of the form:*

$$\mathrm{Sym}^{m_1}(\mathrm{Tr}(S_1)) * \cdots * \mathrm{Sym}^{m_k}(\mathrm{Tr}(S_k))$$

for some $m_1, \ldots, m_k \geqslant 1$ and $2 \leqslant d_1, \ldots, d_k < d$ where $S_j \in \mathrm{CSM}_{d_j}(V')$ for $j \in [k]$ with $m_1 d_1 + \cdots + m_k d_k = d$.

Proof. Our proof is very close to the one by Tao and Ziegler [48]. Where they use the lemma by Bogdanov and Viola [16] to approximate a biased polynomial by its derivatives, we use the newer version of this result by Bhowmick and Lovett [13] that gives bounds independent of field size. By Lemma 3.1 of [13], for any $t \geqslant 1$, we obtain $Q_1, \ldots, Q_c : \mathbb{K}^{nd} \to \mathbb{K}$ and $\Gamma : \mathbb{F}^c \to \mathbb{F}$ where $c = c^{BL}(d, s, t)$ such that:

$$\mathbf{Pr}_{x \in \mathbb{K}^n}[\mathrm{Tr}(T(x)) \neq \Gamma(\mathrm{Tr}(Q_1(x), \ldots, \mathrm{Tr}(Q_c(x))))] \leqslant |\mathbb{K}|^{-t}.$$

Each Q_i here is an additive derivative of T. However, we want an exact representation of $\mathrm{Tr}(T)$ in terms of lower degree polynomials. Kaufmann and Lovett [37] showed that if t is large enough in terms of d, and if $\mathrm{Tr}(Q_1), \ldots, \mathrm{Tr}(Q_c)$ form a *strongly regular* factor in the sense of [37], then in fact, $\mathrm{Tr}(T)$ is exactly a function of $\mathrm{Tr}(Q_1), \ldots, \mathrm{Tr}(Q_c)$. We use the regularization procedure in [12] (Lemma 5.2), which iteratively replaces one of the polynomials in the current collection $\mathrm{Tr}(Q_1), \ldots, \mathrm{Tr}(Q_c)$ with an additive derivative of one of the polynomials in a chosen direction. We do not repeat the definitions and proofs of these results as they closely follow previous work. We also note that we can always write any derivative of a trace of a CSM form in terms of traces of other CSM forms. This follows from the below claim as trace is linear:

▶ **Claim 24.** *Let $s > 0$ be an integer, $L \in \mathrm{CSM}_s(\mathbb{K}^n)$, and $a \in (\mathbb{K}^n)^s$. Then the additive derivative of L in direction a, $D_a L$, can be written as a linear combination of $2^s - 1$ polynomials $(Q_S)_{S \subset [s], S \neq \emptyset}$, where $Q_S \in \mathrm{CSM}_{s-|S|}(\mathbb{K}^n)$ and $c_S \geqslant 1$.*

Proof. We can write

$$D_a L(h_1, \ldots, h_s) = L(h_1 + a_1, \ldots, h_s + a_s) - L(h_1, \ldots, h_s) = \sum_{\substack{S \subset [s] \\ S \neq \emptyset}} L((h_i)_{i \notin S}, (a_i)_{i \in S})$$

where the second equality follows from multilinearity of T. Now letting $Q_S := L((h_i)_{i \notin S}, (1^n)^{s-|S|})$ we have that $Q_S \in \mathrm{CSM}_{s-|S|}(\mathbb{K}^n)$. ◀

The decomposition into concatenation of symmetric powers follows exactly as in [48]. ◀

Therefore, applying Theorem 23 with $T = DP$, we get that for a bounded index subspace V, $\mathrm{Tr}(DP)$ restricted to V^d is a linear combination of a bounded number of expressions of the form:

$$\mathrm{Sym}^{m_1}(\mathrm{Tr}(S_1)) * \cdots * \mathrm{Sym}^{m_k}(\mathrm{Tr}(S_k)) .$$

We next note that since $DP \in \mathrm{CSM}_d(\mathbb{K}^n)$, $\mathrm{Tr}(\alpha_i DP)(h_1, h_2, \ldots, h_d) = \mathrm{Tr}(DP)(\alpha_i h_1, h_2, \ldots, h_d)$. Also, because $S_1, \ldots, S_k$ are each CSM, $S_j(\alpha_i h_1, h_2, \ldots, h_d)$ equals $\alpha_i S_j(h_1, h_2, \ldots, h_d)$ if S_j depends on h_1 and equals $S_j(h_1, h_2, \ldots, h_d)$ otherwise. Hence, for every $i \in [r]$, we get that restricted to the subspace V, $\mathrm{Tr}(\alpha_i DP)$ is a linear combination of a bounded number of "monomials" of the form:

$$\mathrm{Sym}^{m_1}(\mathrm{Tr}(\gamma_{i,1} S_1)) * \cdots * \mathrm{Sym}^{m_k}(\mathrm{Tr}(\gamma_{i,k} S_k)) \tag{4}$$

where $\gamma_{i,1}, \ldots, \gamma_{i,k} \in \mathbb{K}$. Crucially, $S_1, \ldots, S_k$ in all of the above "monomials" are independent of α_i.

Consider any one "monomial" of the form in (4), and we show that there exist non-classical polynomials $Q_1, \ldots, Q_c : \mathbb{K}^n \to \mathbb{T}[Z_1, \ldots, Z_r]$ of weight degrees $< d$ such that

$$\mathrm{Sym}^{m_1}(\mathrm{Tr}(\gamma_{i,1} S_1)) * \cdots * \mathrm{Sym}^{m_k}(\mathrm{Tr}(\gamma_{i,k} S_k)) = D\Delta_i(Q_1, \ldots, Q_k) \tag{5}$$

for some function $\Delta_i : \mathbb{T}[Z_1, \ldots, Z_r]^c \to \mathbb{F}$. Restricted to V^d, our desired form (3) then follows from linearity.

We first show a converse to Lemma 16 which resolves the situation when $m_j = 1$.

▶ **Lemma 25.** *For positive integer d, suppose $S \in \mathrm{CSM}_d(\mathbb{K}^n)$. Then there is a degree-d classical polynomial $Q : \mathbb{K}^n \to \mathbb{K}$ such that $DQ = S$.*

Proof. Tao and Ziegler (Lemma 4.5, [48]) show the same result for CSM forms over $\mathbb{F}^n$, and their proof works without any change. ◀

The next lemma shows that we can integrate symmetric powers of traces of CSM's in terms of non-classical polynomials.

▶ **Lemma 26.** *Let $d \geqslant 2$ and $m \geqslant 1$, $\gamma_1, \ldots, \gamma_r \in \mathbb{K}$, and let $S \in \mathrm{CSM}_d(\mathbb{K}^n)$. Then, there exists a classical polynomial $W : \mathbb{K}^n \to \mathbb{K}$ of weight degree $\leqslant md$ such that $D\mathrm{Tr}(\alpha_i W) = \mathrm{Sym}^m(\mathrm{Tr}(\gamma_i S))$ for all $i \in [r]$. Moreover, if $m \geqslant 2$, then W is a function of a non-classical polynomial of degree $< md$.*

We defer the proof of Lemma 26 to Section 3.1 but we first explain how to complete the proof of Theorem 18. Applying Lemma 26 on each term in the concatenation product in (5), we get for all $j \in [k]$, classical polynomials $W_j : \mathbb{K}^n \to \mathbb{K}$ of weight degree $\leqslant m_j d_j$ such that $D\mathrm{Tr}(\alpha_i W_j) = \mathrm{Sym}^{m_j}(\mathrm{Tr}(\gamma_{i,j} S_j))$, so that the expression in (4) is the derivative polynomial of $U = \prod_{j=1}^{k} \mathrm{Tr}(\alpha_i W_j)$ by Lemma 21. Note that if $k > 1$, then U is already a function of more than one classical polynomial of weight degree $< d$. Otherwise, if $k = 1$, then $m_1 \geqslant 2$ (as $d \geqslant 2$ and $d_1 < d$), and so, U is a function of the non-classical polynomial of degree (and hence, weight degree) $< d$ determining W_1 that is guaranteed to exist by Lemma 26.

We have proved Theorem 18 when all the variables are drawn from V, a subspace of $\mathbb{K}^n$ of co-dimension $t \leqslant t^{23}$. We have shown that we have a degree-d classical polynomial S measurable in non-classical polynomials $\{\tilde{R}_1, \ldots, \tilde{R}_C\}$ of weight degrees $\leqslant d-1$ such that $DP = DS$ on the bounded index subspace V. We can extend the last statement to the subspace $\mathbb{K}^n$ by using a simple derivative trick. Suppose $h'_1, \cdots, h'_t$ are representatives of the quotient group $\mathbb{K}^n/V$. Thus for any $x \in \mathbb{K}^n$, we can have a $i \in [K]$ such that $x - h'_i \in V$. We can write

$$S(x) = S(x - h'_i) - D_{-h'_i} S(x)$$

for x and note that $\deg(D_{-h'_i} S) < d$. This implies that over $\mathbb{K}^n$, S is measurable in $\{D_{-h'_1} S, \ldots, D_{-h'_t} S, \tilde{R}_1, \ldots, \tilde{R}_C : \mathbb{K}^n \to \mathbb{T}[Z_1, \ldots, Z_c]\}$ and $DP = DS$. Now, by letting $Q = S$ and $P' = P - S$, then $DP' = 0$, meaning that weighted degree of P' is less than d, which concludes the proof. ◀

▶ Remark. It is worth noticing that in fact, the proof shows that for every $i \in [r]$, $\mathrm{Tr}(\alpha_i P)$ also has $\mathbb{F}$-rank bounded by $c^{18}(d, s)$, provided that P has degree and weight degree d.

3.1 Proof of Lemma 26

Proof. Our proof follows the outline of the proof of Lemma 6.4 in [48] but with some modifications.

Apply Lemma 25 to get that there is a classical polynomial $R : \mathbb{K}^n \to \mathbb{K}$ of degree d such that $DR = S$. Let $M \geqslant 0$ be an integer such that $p^M \leqslant m < p^{M+1}$. There exists a degree-$(d + M(p-1))$ polynomial $\tilde{R}_M : \mathbb{K}^n \to \mathbb{T}[Z_1, \ldots, Z_r]$ such that $p^M \cdot \tilde{R}_M = \kappa(R)$. And then we pull down the polynomial $\tilde{R}_M$ to the cyclic group $(\mathbb{Z}/p^{M+1}\mathbb{Z})^r$ and we obtain a degree-$(d + M(p-1))$ polynomial $R_M : \mathbb{K}^n \to \left(\mathbb{Z}/p^{M+1}\mathbb{Z}\right)^r$ such that $R_M = R \pmod p$.

Define $\delta_{i,j} = |\mathrm{Tr}(\gamma_i \alpha_j)|$ where $i, j \in [r]$, and thus $\gamma_i = \sum_j \delta_{i,j} \beta_j$. Let

$$W = \sum_{i=1}^{r} \left(\binom{\sum_{j=1}^{r} \delta_{i,j} (R_M)_j}{m} \pmod p \right) \cdot \beta_i$$

where $(R_M)_j$ denotes the jth component of R_M. Define

$$W_{\alpha_i} = \mathrm{Tr}(\alpha_i W) = \binom{\sum_{j=1}^{r} \delta_{i,j}(R_M)_j}{m} \pmod p$$

We claim that

1. $\deg(W_{\alpha_i}) \leqslant md$.
2. $DW_{\alpha_i} = \mathrm{Sym}^m(\gamma_{i,j} DR)$.

Parts 1 and 2 together imply that the weight degree of (W) is at most md. Fix any $i \in [r]$ and define $S = \sum_{j=1}^{r} \delta_{i,j}(R_M)_j$. The last part of the theorem follows from the fact that $d + M(p-1) < md$ when $m > 1$ and that Q can be expressed as a function of $\tilde{R}_M$ which is of degree $d + M(p-1)$. Part 1 will be a special case of the following claim, that is, when $j = 0$ and $m' = m$.

▶ **Lemma 27.** *Let $j \geqslant 0$ and $m' \leqslant m$ be some parameters. Then*

$$\deg\left(\binom{D_{h_1} \cdots D_{h_j} S}{m'} \pmod p\right) \leqslant d - j + (m'-1) \cdot \max(d-j, 1).$$

Proof. We break the claim into two cases:

Case i. $(d-j+(m'-1)\cdot\max(d-j,1) < 0)$: In particular, this implies that $m' < j - d$. We need to show that $\binom{D_{h_1} \cdots D_{h_j} S}{m'}$ is divisible by p. Since $\deg(S) = d + M(p-1)$ then $\deg(D_{h_1} \cdots D_{h_j} S) \leqslant d - j + M(p-1)$ for any $h_1, \ldots, h_j \in \mathbb{K}^n$. And, it is divisible by p^{a+1} whenever $0 \leqslant a \leqslant M$ and $d - j + a(p-1) < 0$. In particular, if we choose $a = \lfloor \frac{m'-1}{p-1} \rfloor$, then $D_{h_1} \cdots D_{h_j} S$ is divisible by $p^{\lfloor \frac{m'-1}{p-1} \rfloor + 1}$. Observe that $\binom{n}{m} \pmod p$ is divisible by p if n is divisible by p^a and $m < p^a$. Since $m' < p^{\lfloor \frac{m'-1}{p-1} \rfloor + 1} = p^{a+1}$, we obtain our claim.

Case ii. $(d-j+(m'-1)\cdot\max(d-j,1) \geqslant 0)$: We will prove this by downward induction on j. The claim is already true for sufficiently large values of j, so we assume inductively that the claim is proven for all larger values of j; and for fixed j, we assume inductively that the claim is proven for all smaller m'. It suffices to show that the expression

$$deg\big(D_{h_{j+1}} \binom{D_{h_1} \cdots D_{h_j} S}{m'} \pmod p\big) \leqslant (d - j + (m'-1) \cdot \max(d-j, 1) - 1$$

holds $\forall h_{j+1} \in \mathbb{K}^n$. We will use the combinatorial identity $\binom{r+s}{m} = \sum_{i=0}^{m} \binom{r}{i}\binom{s}{m-i}$ and then we see that

$$D_h \binom{F}{m} = \sum_{i=1}^{m} \binom{D_h F}{i} \binom{F}{m-i}$$

for $h \in \mathbb{K}^n$ and $F : \mathbb{K}^n \to \mathbb{Z}/p^{M+1}\mathbb{Z}$. We can therefore write

$$D_{h_{j+1}} \binom{D_{h_1} \cdots D_{h_j} S}{m'} = \sum_{i=1}^{m'} \binom{D_{h_1} \cdots D_{h_{j+1}} S}{i} \binom{D_{h_1} \cdots D_{h_j} S}{m'-i}$$

where the both sides of the above equality is over mod p. Now for each summand, we apply the two induction hypothesis and conclude that the degree of the first factor in the right-hand side is $\leqslant d - (j+1) + (i-1)\max(d-(j+1), 1)$ and the degree of the second factor in the

right-hand side is $\leqslant d-j+(m'-i-1)\max(d-j,1)$ and thus the degree of the right-hand side is atmost

$$(d-(j+1)+(i-1)\max(d-(j+1),1))+(d-j+(m'-i-1)\max(d-j,1)) \\ \leqslant d-j+(m'-1)\max(d-j,1)-1$$

whenever $i \geqslant 1$ (by handling the cases $d-j>1$ and $d-j\leqslant 1$ separately), and this concludes the claim that $\deg(W_{\alpha_i}) \leqslant md$. ◀

Now we prove the part 2. We know that

$$D_h\binom{S}{m} = \sum_{i=1}^{m}\binom{D_hS}{i}\binom{S}{m-i}$$

for any $h \in \mathbb{K}^n$. By the above computations, the polynomial $\binom{D_hS}{i}\binom{S}{m-i}$ has degree

$$\leqslant d-1+(i-1)(d-1)+d+(m-i-1)d = md-i.$$

In particular, all the terms of $i>1$ have degree $<mk-1$ and thus will not contribute to $D\binom{S}{m}$. The $i=1$ term can be simplified as $D_h\mathrm{Tr}(\gamma_i R)\binom{S}{m-1}$ by using the equality

$$\mathrm{Tr}(\gamma_i\cdot R) = \mathrm{Tr}\big((\sum_{i=1}^{r}\delta_{i,j}\beta_i)\cdot(\sum_{j=1}^{r}(R)_j\alpha_j)\big) = \sum_{j=1}^{r}\delta_{i,j}\cdot(R)_j$$

We conclude that

$$\begin{aligned}
DW_{\alpha_i}(h_1,\dots,h_{md}) &= D\big(D_{h_{md}}\binom{S}{m}\big)(h_1,\dots,h_{md-1}) \\
&= D\big(\big(D_{h_{md}}\mathrm{Tr}(\gamma_iR)\big)\cdot\binom{S}{m-1}\big)(h_1,\dots,h_{md-1}) \\
&= \big(D\big(D_{h_{md}}\mathrm{Tr}(\gamma_iR)\big) * D\binom{S}{m-1}\big)(h_1,\dots,h_{md-1}) \\
&= \sum_{1\leqslant i_1<\cdots<i_{d-1}<md} D\big(D_{h_{md}}\mathrm{Tr}(\gamma_iR)\big)(h_{i_1},\dots,h_{i_{d-1}},h_{md})\cdot D\binom{S}{m-1}(h_{j_1},\dots,h_{j_{md-d}})
\end{aligned}$$

where $1\leqslant j_1<\cdots<j_{md-d}<md$ are such that $\{j_1,\dots,j_{md-d}\} = \{1,\dots,md-1\}\setminus\{i_1,\dots,i_{d-1}\}$ and by Lemma 21, we have the second equality. Now, by induction on m, we have our claim 2. ◀

4 Equidistribution of regular factors

Our results in this section imply that a high rank collection of polynomials is "as random as possible", subject to the degree and depth bounds of its defining polynomials. We first make some necessary definitions.

4.1 Definitions

A *linear form on k variables* is a vector $L=(w_1,w_2,\dots,w_k)\in\mathbb{K}^k$ that is interpreted as a function from $(\mathbb{K}^n)^k$ to $\mathbb{K}^n$ via the map $(x_1,\dots,x_k)\mapsto w_1x_1+w_2x_2+\cdots+w_kx_k$. A linear form $L=(w_1,w_2,\dots,w_k)$ is said to be *affine* if $w_1=1$. From now, linear forms will always be assumed to be affine. We define wt of a linear form $L=(w_1,\dots,w_k)$ to be $\sum_{i=2}^{k}\mathsf{wt}(w_i)$.

We specify a partial order $\preceq$ among affine forms. We say $(w_1,\dots,w_k)\preceq(w'_1,\dots,w'_k)$ if $|\mathrm{Tr}(\alpha_jw_i)|\leqslant|\mathrm{Tr}(\alpha_jw'_i)|$ for all $i\in[k],j\in[r]$.

▶ **Definition 28** (Affine constraints)**.** An *affine constraint of size m on k variables* is a tuple $A = (L_1, \ldots, L_m)$ of m affine forms $L_1, \ldots, L_m$ over $\mathbb{F}$ on k variables, where: $L_1(x_1, \ldots, x_k) = x_1$. Moreover, it is said to be *weight-closed* if for any affine form L belonging to A, if $L' \preceq L$, then L' also belongs to A.

Observe that a weight-closed affine constraint is of bounded size if and only if all its affine forms are of bounded weight.

Next, we define polynomial factors which play a big role in higher-order Fourier analysis. Here, we restrict ourselves to non-classical polynomials mapping to $\mathbb{T}$ (instead of $\mathbb{T}[Z_1, \ldots, Z_r]$), essentially because throughout we mainly care about the case of constant r.

▶ **Definition 29** (Factor)**.** A *polynomial factor* $\mathcal{B}$ is a sequence of non-classical polynomials $P_1, \ldots, P_C : \mathbb{K}^n \to \mathbb{T}$. We also identify it with the function $\mathcal{B} : \mathbb{K}^n \to \mathbb{T}^C$ mapping x to $(P_1(x), \ldots, P_C(x))$. An *atom* of $\mathcal{B}$ is a preimage $\mathcal{B}^{-1}(y)$ for some $y \in \mathbb{T}^C$. When there is no ambiguity, we will in fact abuse notation and identify an atom of $\mathcal{B}$ with the common value $\mathcal{B}(x)$ of all x in the atom.

The *partition induced by* $\mathcal{B}$ is the partition of $\mathbb{K}^n$ given by $\left\{\mathcal{B}^{-1}(y) : y \in \mathbb{T}^C\right\}$. The *complexity* of $\mathcal{B}$, denoted $|\mathcal{B}|$, is the number of defining polynomials C. The *order* of $\mathcal{B}$, denoted $\|\mathcal{B}\|$, is the total number of atoms in $\mathcal{B}$. The *degree* of $\mathcal{B}$ is the maximum degree among its defining polynomials $P_1, \ldots, P_C$.

Note that due to Definition 9, if $\mathcal{B}$ is defined by polynomials $P_1, \ldots, P_C$,

$$\|\mathcal{B}\| = \prod_{i=1}^{C} p^{\mathrm{depth}(P_i)+1}.$$

From henceforth, since we will work only with non-classical polynomials $P : \mathbb{K}^n \to \mathbb{T}$, rank will denote their $\mathbb{F}$-rank.

▶ **Definition 30** (Rank and Regularity of Polynomial Factor)**.** Let $\mathcal{B}$ be a polynomial factor defined by the sequence $P_1, \ldots, P_c : \mathbb{K}^n \to \mathbb{T}$ with respective depths $k_1, \ldots, k_c$. Then, the rank of $\mathcal{B}$ is $\min_{(a_1,\ldots,a_c)} \mathrm{rank}(\sum_{i=1}^{c} a_i P_i)$ where the minimum is over $(a_1, \ldots, a_c) \in \mathbb{Z}^c$ such that $(a_1 \mod p^{k_1+1}, \ldots, a_c \mod p^{k_c+1}) \neq (0, \ldots, 0)$.

Given a polynomial factor $\mathcal{B}$ and a non decreasing function $r : \mathbb{Z}^+ \to \mathbb{Z}^+$, $\mathcal{B}$ is r-regular if $\mathcal{B}$ is of rank at least $r(|\mathcal{B}|)$.

▶ **Definition 31** (Semantic and Syntactic refinement)**.** Let $\mathcal{B}$ and $\mathcal{B}'$ be polynomial factors. A factor $\mathcal{B}'$ is a *syntactic refinement* of $\mathcal{B}$, denoted by $\mathcal{B}' \succeq_{syn} \mathcal{B}$ if the set of polynomials defining $\mathcal{B}$ is a subset of the set of polynomials defining $\mathcal{B}'$. It is a *semantic refinement*, denoted by $\mathcal{B}' \succeq_{sem} \mathcal{B}$ if for every $x, y \in \mathbb{K}^n$, $\mathcal{B}'(x) = \mathcal{B}'(y)$ implies $\mathcal{B}(x) = \mathcal{B}(y)$. Clearly, a syntactic refinement is also a semantic refinement.

Our next lemma is the workhorse that allows us to convert any factor into a regular one.

▶ **Lemma 32** (Polynomial Regularity Lemma)**.** *Let $r : \mathbb{Z}^+ \to \mathbb{Z}^+$ be a non-decreasing function and $d > 0$ be an integer. Then, there is a function $C_{32}^{(r,d)} : \mathbb{Z}^+ \to \mathbb{Z}^+$ such that the following is true. Suppose $\mathcal{B}$ is a factor defined by polynomials $P_1, \ldots, P_C : \mathbb{K}^n \to \mathbb{T}$ of additive degree at most d. Then, there is an r-regular factor $\mathcal{B}'$ consisting of polynomials $Q_1, \ldots, Q_{C'} : \mathbb{K}^n \to \mathbb{T}$ of additive degree $\leqslant d$ such that $\mathcal{B}' \succeq_{sem} \mathcal{B}$ and $C' \leqslant C_{32}^{(r,d)}(C)$.*

Moreover, if $\mathcal{B}$ is itself a refinement of some polynomial factor $\hat{\mathcal{B}}$ that has rank $> (r(C') + C')$, then additionally $\mathcal{B}'$ will be a syntactic refinement of $\hat{\mathcal{B}}$.

Proof. Follows directly from Lemma 2.18 of [9] by identifying $\mathbb{K}^n$ with $\mathbb{F}^{rn}$. ◀

In fact, the regularization process of Lemma 32 can be implemented in time $O(n^{d+1})$ [12].

Finally, we'll use the Gowers inverse theorem proved by Tao and Ziegler [48] for non-classical polynomials mapping to $\mathbb{T}$.

▶ **Theorem 33** (Theorem 1.20 of [48]). *Suppose $\delta > 0$ and $d \geqslant 1$ is an integer. There exists an $r = r_{33}(\delta, d)$ such that the following holds. If a non-classical polynomial $P : \mathbb{K}^n \to \mathbb{T}$ with degree d satisfies $\|P\|_{U^d} \geqslant \delta$, then* $\operatorname{rank}(P) \leqslant r$.

4.2 Equidistribution results

Let us start with the following simple observation.

▶ **Lemma 34.** *Given $\varepsilon > 0$, let $\mathcal{B}$ be a polynomial factor of degree $d > 0$, complexity C and rank $r_{34}(d, \varepsilon)$, defined by a sequence of non-classical polynomials $P_1, \ldots, P_C : \mathbb{K}^n \to \mathbb{T}$ having respective depths $k_1, \ldots, k_C$. Suppose $\alpha = (\alpha_1, \ldots, \alpha_C) \in \mathbb{U}_{k_1+1} \times \cdots \times \mathbb{U}_{k_C+1}$. Then:*

$$\mathbf{Pr}_x[\mathcal{B}(x) = \alpha] = \frac{1}{\|\mathcal{B}\|} \pm \varepsilon.$$

Proof. This is standard. See for example lemma 3.2 of [9]. ◀

In our applications though, we will often need not just $\mathcal{B}(x)$ to be nearly uniformly distributed but the tuple $(\mathcal{B}(x) : x \in H)$ for a set $H \subseteq \mathbb{K}^n$ to be nearly uniformly distributed. In particular, we consider the case when H is an affine subspace of $\mathbb{K}^n$. The following lemma is key.

▶ **Lemma 35** (Near orthogonality). *Let $A = (L_1, \ldots, L_m)$ be a weight-closed affine constraint of bounded size on ℓ variables. Suppose $\mathcal{B}$ is a polynomial factor of degree d and rank $\geqslant r^{(33)}(d, \delta)$, defined by the sequence of non-classical polynomials $P_1, \ldots, P_c : \mathbb{K}^n \to \mathbb{T}$. Let $\Lambda = (\lambda_{ij})_{i \in [c], j \in [m]}$ be a tuple of integers. Define:*

$$P_\Lambda(x_1, \ldots, x_k) = \sum_{i \in [c], j \in [m]} \lambda_{ij} P_i(L_j(x_1, \ldots, x_\ell)).$$

Then one of the following is true.

1. *For every $i \in [c]$, it holds that $\sum_{j \in [m]} \lambda_{ij} Q_i(L_j(\cdot)) \equiv 0$ for all polynomials $Q_i : \mathbb{K}^n \to \mathbb{T}$ with the same degree and depth as P_i. Clearly, this implies $P_\Lambda \equiv 0$.*
2. *$P_\Lambda \not\equiv 0$. Moreover,* $\operatorname{bias}(P_\Lambda) \leqslant \delta$.

Proof. For $j \in [m]$, let $(w_{j,1}, \ldots, w_{j,\ell}) \in \mathbb{K}^\ell$ denote the affine form given by L_j. Note that $w_{j,1} = 1$.

For each i, we do the following. If for some j, we have $\mathsf{wt}(L_j) > \deg(\lambda_{i,j} P_i)$, $\lambda_{i,j} \neq 0$, then, $L_j(x_1, \ldots, x_\ell) = x_1 + \sum_{i=2}^{\ell} \left(\sum_{k=1}^{r} u_{i,k} \cdot \beta_k\right) x_i$ where $\boldsymbol{\beta}$ is the dual basis to $\boldsymbol{\alpha}$, each $u_{i,k} \in [0, p-1]$ and $\sum_{i,k} u_{i,k} > \deg(\lambda_{i,j} P_i)$. Using Definition 2, we can replace $\lambda_{i,j} P_i(L_j)$ by a $\mathbb{Z}$-linear combination of $P_i(L_{j'})$ where $L_{j'} \preceq L_j$ until no such j exists. This is where we use the fact that the affine constraint is weight-closed. Suppose the new coefficients are denoted by $(\lambda'_{i,j})$. If the $\lambda'_{i,j}$ are all zero, then for every $i \in [c]$ individually, $\sum_{j \in [m]} P_i(L_j(x_1, \ldots, x_\ell)) \equiv 0$. Indeed, $\sum_{j \in [m]} Q_i(L_j(x_1, \ldots, x_\ell)) \equiv 0$ for any Q_i with the same degree and depth, as the transformation from $\lambda_{i,j}$ to $\lambda'_{i,j}$ did not use any other information about P_i.

Else some $\lambda'_{i,j} \neq 0$. Also, $\mathsf{wt}_{\boldsymbol{\alpha}}(L_j) \leqslant \deg(\lambda'_{i,j} P_i)$. Then we show the second part of the lemma, that is $|\mathbb{E}[e(P_\Lambda(x_1, \ldots, x_k)]| \leqslant \delta$.

Suppose without loss of generality that the following is true.

- $\lambda'_{i,1} \neq 0$ for some $i \in [C]$.
- L_1 is maximal in the sense that for every $j \neq 1$, either $\lambda'_{i,j} = 0$ for all $i \in [C]$ or $\mathsf{wt}_{\boldsymbol{\alpha}}(w_{j,s}) < \mathsf{wt}_{\boldsymbol{\alpha}}(w_{1,s})$ for some $s \in [\ell]$.

For $a = (a_1, \ldots, a_\ell) \in \mathbb{K}^\ell$ and $y \in \mathbb{K}^n$ and $P : \mathbb{K}^n \to \mathbb{T}$, define

$$\overline{D}_{a,y} P(x_1, \ldots, x_\ell) = P(x_1 + a_1 y, \ldots, x_\ell + a_\ell y) - P(x_1, \ldots, x_\ell).$$

Then

$$\overline{D}_{a,y}(P_i \circ L_j)(x_1, \ldots, x_\ell) = (D_{L_j(a)y} P_i)(L_j(x_1, \ldots, x_\ell)).$$

Let $\Delta = \mathsf{wt}_\alpha(L_1) \leqslant d$. Define $a_1, \ldots, a_\Delta$ be the set of vectors of the form $(-w, 0, \ldots, 1, 0, \ldots, 0)$ where 1 is in the ith coordinate for $i \in [2, \ell]$ and for all $w \in \mathbb{K}$ satisfying $0 \leqslant \mathsf{wt}_{\boldsymbol{\alpha}}(w) < \mathsf{wt}_{\boldsymbol{\alpha}}(w_{1,i})$. Note that $\langle L_1, a_k \rangle \neq 0$ for $k \in [\Delta]$ but for any $j > 1$ there exists some $k \in [\Delta]$ such that $\langle L_j, a_k \rangle = 0$. Thus,

$$\mathop{\mathbb{E}}_{y_1,\ldots,y_\Delta,x_1,\ldots,x_\ell} \left[e\left((\overline{D}_{a_\Delta,y_\Delta} \ldots \overline{D}_{a_1.y_1} P_\Lambda)(x_1, \ldots, x_\ell) \right) \right] = \left\| \sum_{i=1}^{C} \lambda'_{i,1} P_i \right\|_{U^\Delta}^{2^\Delta}.$$

The rest of the analysis is same as Theorem 3.3 in [9] and we skip it here. ◀

We can now use Lemma 35 to prove our result on equidistribution of regular factors over affine subspaces of $\mathbb{K}^n$.

▶ **Theorem 36.** *Let $\varepsilon > 0$. Let $\mathcal{B}$ be a polynomial factor defined by non-classical polynomials $P_1, \ldots, P_c : \mathbb{K}^n \to \mathbb{T}$ with respective degrees $d_1, \ldots, d_c \in \mathbb{Z}^+$ and depths $k_1, \ldots, k_c \in \mathbb{Z}^{\geqslant 0}$. Suppose $\mathcal{B}$ has rank at least $r^{(33)}(d, \varepsilon)$ where $d = \max(d_1, \ldots, d_c)$. Let $A = (L_1, \ldots, L_m)$ be a weight-closed affine constraint. For every $i \in [c]$, define Λ_i to be the set of tuples $(\lambda_1, \ldots, \lambda_m) \in [0, p^{k_i+1} - 1]$ such that $\sum_{j=1}^m \lambda_j Q_i(L_j(\cdot)) \equiv 0$ for all non-classical polynomials Q_i with the same degree and depth as P_i.*

Consider $(\alpha_{i,j} : i \in [c], j \in [m]) \in \mathbb{T}^{cm}$ such that for every $i \in [c]$ and for every $(\lambda_1, \ldots, \lambda_m) \in \Lambda_i$, $\sum_{j=1}^m \lambda_j \alpha_{i,j} = 0$. Then:

$$\mathbf{Pr}_{x_1,\ldots,x_\ell \in \mathbb{K}^n}[\mathcal{B}(L_j(x_1, \ldots, x_\ell)) = (\alpha_{1,j}, \ldots, \alpha_{c,j}) \ \forall j \in [m]] = \frac{\prod_{i=1}^c |\Lambda_i|}{\|\mathcal{B}\|^m} \pm \varepsilon$$

Proof.

$$\begin{aligned}
&\mathbf{Pr}_{x_1,\ldots,x_\ell \in \mathbb{K}^n}[\mathcal{B}(L_j(x_1, \ldots, x_\ell)) = (\alpha_{1,j}, \ldots, \alpha_{c,j}) \ \forall j \in [m]] \\
&= \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell} \left[\prod_{i,j} \frac{1}{p^{k_i+1}} \sum_{\lambda_{i,j}=0}^{p^{k_i+1}-1} e(\lambda_{i,j}(P_i(L_j(x_1, \ldots, x_\ell)) - \alpha_{i,j})) \right] \\
&= \left(\prod_i p^{-(k_i+1)} \right)^m \sum_{\in \prod_{i,j}[0,p^{k_i+1}-1]}^{(\lambda_{i,j})} e\left(-\sum_{i,j} \lambda_{i,j}\alpha_{i,j} \right) \mathbb{E}\left[e\left(\sum_{i,j} \lambda_{i,j} P_i(L_j(x_1, \ldots, x_\ell)) \right) \right] \\
&= p^{-m\sum_{i=1}^c (k_i+1)} \cdot \left(\prod_{i=1}^c |\Lambda_i| \ \pm \ \varepsilon p^{m\sum_{i=1}^c (k_i+1)} \right)
\end{aligned}$$

The last line is due to the observation that from Lemma 35, $\sum_{i=1}^c \sum_{j=1}^m \lambda_{i,j} P_i(L_j(x_1, \ldots, x_\ell)) \equiv 0$ if and only if for every $i \in [c]$, $(\lambda_{i,1}, \ldots, \lambda_{i,m}) \in \Lambda_i \pmod{p^{k_i+1}}$. So, $\sum_{i,j} \lambda_{i,j} P_i(L_j(\cdot))$ is identically 0 for $\prod_i |\Lambda_i|$ many tuples $(\lambda_{i,j})$ and for those tupes, $\sum_{i,j} \lambda_{i,j}\alpha_{i,j} = 0$ also. ◀

Note that in Theorem 36, if ε is a constant, m needs to be bounded for the claim to be non-trivial, which in turn requires that the affine forms in L be of bounded weight.

4.3 Preservation of Locally Characterized Properties

4.3.1 Local Characterization

As described in the introduction, by a locally characterized property, we informally mean a property for which non-membership can be certified by a finite sized witness. Specifically for affine-invariant properties, we define:

▶ **Definition 37** (Locally characterized properties)**.**

- An *induced affine constraint of size m on ℓ variables* is a pair (A, σ) where A is an affine constraint of size m on ℓ variables and $\sigma \in [R]^m$.
- Given such an induced affine constraint (A, σ), a function $f : \mathbb{K}^n \to [R]$ is said to be (A, σ)-*free* if there exist no $x_1, \ldots, x_\ell \in \mathbb{K}^n$ such that $(f(L_1(x_1, \ldots, x_\ell)), \ldots, f(L_m(x_1, \ldots, x_\ell))) = \sigma$. On the other hand, if such $x_1, \ldots, x_\ell$ exist, we say that f *induces* (A, σ) *at* $x_1, \ldots, x_\ell$.
- Given a (possibly infinite) collection $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^i, \sigma^i), \ldots\}$ of induced affine constraints, a function $f : \mathbb{K}^n \to [R]$ is said to be $\mathcal{A}$-*free* if it is (A^i, σ^i)-free for every $i \geqslant 1$. The size of $\mathcal{A}$ is the size of the largest induced affine constraint in $\mathcal{A}$.
- Additionally, $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^K, \sigma^K)\}$ is a W-light affine system if there exists a basis $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)$ such that $\mathsf{wt}_{\boldsymbol{\alpha}}(A^i) \leqslant W$ for all $i \in [K]$.
- A property $\mathcal{P} \subseteq \{\mathbb{K}^n \to [R]\}$ is said to be K, W*-lightly locally characterized* if it is equivalent to $\mathcal{A}$-freeness for some W-light affine system $\mathcal{A}$ whose size is $\leqslant K$.

We recall that Kaufman and Ron [38] show that:

▶ **Theorem 38** ([38])**.** *The property $\mathcal{P}_d = \{P : \mathbb{K}^n \to \mathbb{K} : \deg(P) \leqslant d\}$ is $q^{\lceil (d+1)/(q-q/p) \rceil}$, pr $\lceil (d+1)/(q-q/p) \rceil$-lightly locally characterized.*

4.3.2 Main Result on Property Preservation

▶ **Theorem 39.** *Let $\mathcal{P} \subset \{\mathbb{K}^n \to \mathbb{K}\}$ be a K, W-lightly locally characterized property. For an integer d, suppose $P_1, \ldots, P_c : \mathbb{K}^n \to \mathbb{T}$ are polynomials of additive degree $\leqslant d$, forming a factor of rank $> r_{39}(d, K)$, and $\Gamma : \mathbb{T}^c \to \mathbb{K}$ is a function such that $F : \mathbb{K}^n \to \mathbb{K}$ defined by $F(x) = \Gamma(P_1(x), \ldots, P_c(x))$ satisfies $\mathcal{P}$.*

For every collection of additive polynomials $Q_1, \ldots, Q_c : \mathbb{K}^n \to \mathbb{T}$ with $\deg(Q_i) \leqslant \deg(P_i)$ and $\mathrm{depth}(Q_i) \leqslant \mathrm{depth}(P_i)$ for all $i \in [c]$, if $G : \mathbb{K}^n \to \mathbb{K}$ is defined by $G(x) = \Gamma(Q_1(x), \ldots, Q_c(x))$, then $G \in \mathcal{P}$ too.

Proof. For the sake of contradiction, suppose $G \notin \mathcal{P}$. Then, for a weight-closed affine constraint consisting of K' linear forms $L_1, \ldots, L_{K'}$, there exist $x_1, \ldots, x_\ell$ such that $(G(L_1(x_1, \ldots, x_\ell)), \ldots, G(L_{K'}(x_1, \ldots, x_\ell)))$ which form a witness to $G \notin \mathcal{P}$. Note that K' is a function of only K and W because the affine forms characterizing $\mathcal{P}$ can be made weight $\leqslant W$ by a choice of basis for $\mathbb{K}$ over $\mathbb{F}$ and then completed into a weight-closed constraint. So, there exists $x_1, \ldots, x_\ell \in \mathbb{K}^n$ such that the tuple $B = (Q_i(L_j(x_1, \ldots, x_\ell)) : j \in [K'], i \in [c]) \in \mathbb{T}^{cK'}$ is a proof of the fact that $G \notin \mathcal{P}$.

Now we argue that there exist $x'_1, \ldots, x'_\ell$ such that $(P_i(L_j(x'_1, \ldots, x'_\ell)) : i \in [c], j \in [K])$ equals B, thus showing that $F \notin P$, a contradiction. Notice that B satisfies the conditions required of α in Theorem 36. So by Theorem 36,

$$\mathbf{Pr}_{x'_1, \ldots, x'_\ell} [(P_i(L_j(x'_1, \ldots, x'_\ell) : i \in [c], j \in [K]) = B] > 0$$

if the rank of the factor formed by $P_1, \ldots, P_c$ is more than $r^{(33)}\left(d, \frac{1}{2\|\mathcal{B}\|^K}\right)$, where $\|\mathcal{B}\| = p^{\sum_{i=1}^{c}(\text{depth}(P_i)+1)}$. ◀

In our applications, we will use Theorem 39 for the property of having bounded degree, which is lightly locally characterized by Theorem 38.

5 List decoding of RM codes

We state the following corollary which we need in the proof to follow. We only state a special case of it which is enough.

▶ **Corollary 40** (Corollary 3.3 of [14]). *Let $g : K \to K$, $\varepsilon > 0$. Then there exist $c \leqslant 1/\varepsilon^2$ functions $h_1, h_2, \ldots, h_c \in \text{RM}_{\mathbb{K}}(n, d)$ such that for every $f \in \text{RM}_{\mathbb{K}}(n, d)$, there is a function $\Gamma_f : \mathbb{K}^c \to \mathbb{K}$ such that*

$$\mathbf{Pr}_x[\Gamma_f(h_1(x), \ldots, h_c(x)) = f(x)] \geqslant \mathbf{Pr}_x[g(x) = f(x)] - \varepsilon.$$

▶ **Theorem 2** (restated). *Let $\mathbb{K} = \mathbb{F}_q$ be an arbitrary finite field. Let $\varepsilon > 0$ and $d, n \in \mathbb{N}$. Then,*

$$\ell_{\mathbb{K}}(d, n, \delta_{\mathbb{K}}(d) - \varepsilon) \leqslant c_{q,d,\varepsilon}.$$

Proof. We follow the proof structure in [14]. Let $g : \mathbb{K}^n \to \mathbb{K}$ be a received word. Apply Corollary 40 with approximation parameter $\varepsilon/2$ gives $\mathcal{H}_0 = \{h_1, \ldots, h_c\} \subseteq \text{RM}_{\mathbb{K}}(n, d)$, $c \leqslant 4/\varepsilon^2$ such that, for every $f \in \text{RM}_{\mathbb{K}}(n, d)$, there is a function $\Gamma_f : \mathbb{K}^c \to \mathbb{K}$ satisfying

$$\mathbf{Pr}[\Gamma_f(h_1(x), h_2(x), \ldots, h_c(x)) = f(x)] \geqslant \mathbf{Pr}[g(x) = f(x)] - \varepsilon/2.$$

B

$$\mathbf{Pr}[\Gamma'_f(\text{Tr}(\alpha_i h_j(x)) : 1 \leqslant i \leqslant r, 1 \leqslant j \leqslant c) = F(\text{Tr}(\alpha_i f(x)) : 1 \leqslant i \leqslant r)] \geqslant d/q + \varepsilon/2,$$

where $\Gamma'_f : \mathbb{F} \to \mathbb{K}$ and $F : \mathbb{F}^r \to \mathbb{K}$. From here onwards, we identify $\mathbb{F}$ with $\mathbb{U}_1$. Let $\mathcal{H} = \{\text{Tr}(\alpha_i h_j(x)) : 1 \leqslant i \leqslant r, 1 \leqslant j \leqslant c\}$ and $\mathcal{H}_F = \{\text{Tr}(\alpha_i f(x)) : 1 \leqslant i \leqslant r)\}$.

Let $r_1, r_2 : \mathbb{N} \to \mathbb{N}$ be two non decreasing functions to be specified later, and let $C^{(32)}_{r,d}$ be as given in Lemma 32. We will require that for all $m \geqslant 1$,

$$r_1(m) \geqslant r_2(C^{(32)}_{r_2,d}(m+1)) + C^{(32)}_{r_2,d}(m+1) + 1. \tag{6}$$

As a first step, we r_1-regularize $\mathcal{H}$ by Lemma 32. This gives an r_1-regular factor $\mathcal{B}'$ of degree at most d, defined by polynomials $H_1, \ldots, H_c : \mathbb{K}^n \to \mathbb{T}$, $c' \leqslant C^{(32)}_{r_1,d}(cr)$ and $\text{rank}(\mathcal{B}') \geqslant r_1(c')$. We denote $\mathcal{H}' = \{H_1, \ldots, H_{c'}\}$. Let $\text{depth}(H_i) = k_i$ for $i \in [c']$. Let $G_f : \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1} \to \mathbb{U}_1$ be defined such that

$$\Gamma_f(h_1(x), \ldots, h_c(x)) = G_f(h'_1(x), \ldots, h'_{c'}(x)).$$

Next, given any polynomial $f : \mathbb{K}^n \to \mathbb{K}$ of degree at most d, we will show that if $\mathbf{Pr}[f(x) \neq g(x)] \leqslant \delta(d) - \varepsilon$, then f is measurable with respect to $\mathcal{H}'$ and this would upper bound the number of such polynomials by $c'(q, d, \varepsilon)$ independent on n.

Fix such a polynomial f. Call $F_i = \text{Tr}(\alpha_i f)$. Appealing again to Lemma 32, we r_2-regularize $\mathcal{B}_f := \mathcal{B}' \bigcup \mathcal{H}_F$. We get an r_2-regular factor $\mathcal{B}'' \succeq_{syn} \mathcal{B}'$ defined by the collection

$\mathcal{H}'' = \{H_1, \ldots, H_{c'}, H'_1, \ldots, H'_{c''}\}$. Note that it is a syntactic refinement of $\mathcal{B}'$ as by our choice of r_1,

$$\operatorname{rank}(\mathcal{B}') \geqslant r_1(c') \geqslant r_2(C^{(32)}_{r_2,d}(c'+1)) + C^{(32)}_{r_2,d}(c'+1) + 1 \geqslant r_2(|\mathcal{B}''|) + |\mathcal{B}''| + 1.$$

We will choose r_2 such that for all $m \geqslant 1$,

$$r_2(m) = \max\left(r^{(34)}_d\left(\frac{\varepsilon/4}{\left(p^{\lfloor \frac{d-1}{p-1} \rfloor + 1}\right)^m}\right), r^{(39)}_d(m)\right). \tag{7}$$

Since each F_i is measurable with respect to $\mathcal{B}''$, there exists $F' : S \to \mathbb{U}_1$ such that

$$f(x) = F'(H_1(x), \ldots, H_{c'}(x), H'_1(x), \ldots, H'_{c''}(x)).$$

Summing up, we have

$$\mathbf{Pr}[G(H_1(x), H_2(x), \ldots, H_{c'}(x)) = F'(H_1(x), \ldots, H_{c'}(x), H'_1(x), \ldots, H'_{c''}(x))] \geqslant d/q + \varepsilon/2.$$

We next show that we can have each polynomial in the factor have a disjoint set of inputs. This would simplify the analysis considerably.

▶ **Claim 41.** *Let x^i, y^j, $i \in [c'], j \in [c'']$ be pairwise disjoint sets of $n \in \mathbb{N}$ variables each. Let $n' = n(c' + c'')$. Let $\tilde{f} : \mathbb{K}^{n'} \to \mathbb{K}$ and $\tilde{g} : \mathbb{K}^{n'} \to \mathbb{K}$ be defined as*

$$\tilde{f}(x) = F(H_1(x^1), \ldots, H_{c'}(x^{c'}), H'_1(y^1), \ldots, H'_{c''}(y^{c''}))$$

and

$$\tilde{g}(x) = G(H'_1(x^1), \ldots, H_{c'}(x^{c'})).$$

Then $\deg(\tilde{f}) \leqslant d$ *and*

$$\left|\mathbf{Pr}_{x \in \mathbb{F}^{n'}}[\tilde{f}(x) = \tilde{g}(x)] - \mathbf{Pr}_{x \in \mathbb{F}^n}[f(x) = G_f(h'_1(x), h'_2(x), \ldots, h'_c(x))]\right| \leqslant \varepsilon/4.$$

Proof. The bound $\deg(\tilde{f}) \leqslant \deg(f) \leqslant d$ follows from Lemma 39. To establish the bound on $\mathbf{Pr}[\tilde{f} = \tilde{g}]$, for each $s \in S$ let

$$p_1(s) = \mathbf{Pr}_{x \in \mathbb{F}^n}[(h'_1(x), \ldots, h'_{c'}(x), h''_1(x), \ldots, h''_{c''}(x)) = s].$$

Applying Lemma 34 and since our choice of r_2 satisfies $\operatorname{rank}(\mathcal{H}'') \geqslant r^{(34)}_d(\varepsilon/4|S|)$, we have that p_1 is nearly uniform over S,

$$p_1(s) = \frac{1 \pm \varepsilon/4}{|S|}.$$

Similarly, let

$$p_2(s) = \mathbf{Pr}_{x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''} \in \mathbb{F}^n}[(h'_1(x^1), \ldots, h'_{c'}(x^{c'}), h''_1(y^1), \ldots, h''_{c''}(y^{c''})) = s].$$

Note that the rank of the collection of polynomials $\{h'_1(x^1), \ldots, h'_{c'}(x^{c'}), h''_1(y^1), \ldots, h''_{c''}(y^{c''})\}$ defined over $\mathbb{F}^{n'}$ cannot be lower than that of $\mathcal{H}''$. Applying Lemma 34 again gives

$$p_2(s) = \frac{1 \pm \varepsilon/4}{|S|}.$$

For $s \in S$, let $s' \in \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1}$ be the restriction of s to first c' coordinates, that is, $s' = (s_1, \ldots, s_{c'})$. Thus

$$\begin{aligned}\mathbf{Pr}_{x\in\mathbb{F}^{n'}}[\tilde{f}(x) = \tilde{g}(x)] &= \sum_{s\in S} p_2(s) 1_{F(s)=G_f(s')} \\ &= \sum_{s\in S} p_1(s) 1_{F(s)=G_f(s')} \pm \varepsilon/4 \\ &= \mathbf{Pr}_{x\in\mathbb{F}^n}[f(x) = G_f(h'_1(x), h'_2(x), \ldots, h'_c(x))] \pm \varepsilon/4.\end{aligned}$$

◀

So, we obtain that

$$\mathbf{Pr}_{x\in\mathbb{F}^{n'}}[\tilde{f}(x) = \tilde{g}(x)] \geqslant \mathbf{Pr}_{x\in\mathbb{F}^n}[f(x) = G_f(h'_1(x), \ldots, h'_{c'}(x))] - \varepsilon/4 \geqslant 1 - \delta(d) + \varepsilon/4.$$

Next, we need the following variant of the Schwartz-Zippel lemma from [14].

▶ **Claim 42.** *Let $d, n_1, n_2 \in \mathbb{N}$. Let $f_1 : \mathbb{K}^{n_1+n_2} \to \mathbb{K}$ and $f_2 : \mathbb{K}^{n_1} \to \mathbb{K}$ be such that* $\deg(f_1) \leqslant d$ *and*

$$\mathbf{Pr}[f_1(x_1, \ldots, x_{n_1+n_2}) = f_2(x_1, \ldots, x_{n_1})] > 1 - \delta(d)$$

Then, f_1 does not depend on $x_{n_1+1}, \ldots, x_{n_1+n_2}$.

With Claim 42 applied to $f_1 = \tilde{f}, f_2 = \tilde{g}, n_1 = nc', n_2 = nc''$. We obtain that $\tilde{f}$ does not depend on $y^1, \ldots, y^{c''}$. Hence,

$$\tilde{f}(x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''}) = F(H'_1(x^1), \ldots, H'_{c'}(x^{c'}), C_1, \ldots, C_{c''})$$

where $C_j = H''_j(0)$ for $j \in [c'']$. If we substitute $x^1 = \ldots = x^{c'} = x$ we get that

$$f(x) = F(H'_1(x), \ldots, H'_{c'}(x), H''_1(x), \ldots, H''_{c''}(x)) = F(H'_1(x), \ldots, H'_{c'}(x), C_1, \ldots, C_{c''}),$$

which shows that f is measurable with respect to $\mathcal{H}'$, as claimed. ◀

6 Polynomial decomposition

We first formally define the problem for which we claim a polynomial time algorithm.

▶ **Definition 43.** Given $k \in \mathbb{N}$ and $\Delta = (\Delta_1, \ldots, \Delta_k) \in \mathbb{N}^k$ and a function $\Gamma : \mathbb{K}^k \to \mathbb{K}$, a function $P : \mathbb{K}^n \to \mathbb{K}$ is (k, Δ, Γ)-structured if there exist polynomials $P_1, \ldots, P_k : \mathbb{K}^n \to \mathbb{K}$ with $\deg(P_i) \leqslant \Delta_i$ such that for $x \in \mathbb{K}^n$, we have

$$P(x) = \Gamma(P_1(x), \ldots, P_k(x)).$$

The polynomials $P_1, \ldots, P_k$ form a (k, Δ, Γ)-decomposition.

The main result we prove is the following.

▶ **Theorem 44.** *Let $k \in \mathbb{N}$. For every $\Delta = (\Delta_1, \ldots, \Delta_k) \in \mathbb{N}^k$ and every function $\Gamma : \mathbb{K}^k \to \mathbb{K}$, there is a randomized algorithm A that on input $P : \mathbb{K}^n \to \mathbb{K}$ of degree d, runs in time $\mathrm{poly}_{q,k,\Delta}(n^{d+1})$ and outputs a (k, Δ, Γ)-decomposition of P if one exists while otherwise returning NO.*

We first show that the notion of rank is robust to hyperplane restrictions over nonprime fields. More precisely, we have the following.

▶ **Lemma 45.** *Let $P : \mathbb{K}^n \to \mathbb{T}$ be a non-classical polynomial such that* $\mathrm{rank}(P) \geqslant r$*. Let H be a hyperplane in $\mathbb{K}^n$. Then the restriction of P to H has rank at least $r - q$.*

Proof. Without loss of generality, let H be defined by $x_1 = 0$. Let $P' : \mathbb{K}^{n-1} \to \mathbb{T}$ be the restriction of P defined by $P'(y) = P(0y)$. Let $\pi : \mathbb{K}^n \to \mathbb{K}^{n-1}$ be the map $\pi(x_1x_2 \dots x_n) = x_2 \dots x_n$. Let $P'' : \mathbb{K}^n \to \mathbb{T}$ be defined by $P''(x) = P(x) - P' \circ \pi$. Then $P''(x) = 0$ for $x \in H$. For $i \in \mathbb{K} \setminus \{0\}$, let $h_i = (i, 0, \dots, 0)$. Then, for $y \in H$, define $R_j : \mathbb{K}^n \to \mathbb{T}$ by

$$R_j(y) = P''(y + h_j) = (D_{h_j}P'')(y).$$

Note that $\deg(R_j) \leqslant d - 1$. Now, since $P(x) = P''(x) + P' \circ \pi(x)$, we have

$$P(x) = \Gamma(P' \circ \pi, x_1, \{R_y(x) : y \in \mathbb{F}\}).$$

Now, if $\mathrm{rank}(P') \leqslant r$, then $\mathrm{rank}(P' \circ \pi) \leqslant r$ and hence $\mathrm{rank}(P) \leqslant r + q$. This finishes the proof. ◀

We now start with the proof of Theorem 44.

Proof. Let $R_1 : \mathbb{N} \to \mathbb{N}$ be defined as $R_1(m) = R_2(c_{32}^{(R_1,d)}(m+k)) + c_{32}^{(R_1,d)}(m+k) + q$ where $R_2 : \mathbb{N} \to \mathbb{N}$ will be specified later.

We have that $P(x) = \sum_i \beta_i \mathrm{Tr}(\alpha_i P(x))$ for the dual basis $\beta_1, \dots, \beta_r$. Set $f_i(x) = \mathrm{Tr}(\alpha_i P(x))$. Identifying $\mathbb{F}$ with $\mathbb{U}_1$ we treat $f_i : \mathbb{K}^n \to \mathbb{T}$. Regularize $\{f_1, \dots, f_r\}$ using the algorithm of [12] to find R_1-regular $\mathcal{B} = \{g_1, \dots, g_C : \mathbb{K}^n \to \mathbb{T}\}$ where $C \leqslant c_{32}^{(R_1,d)}(r)$. So, $f_i(x) = G_i(g_1(x), \dots, g_C(x))$ and $P(x) = \sum_i \alpha_i G_i(g_1(x), \dots, g_C(x))$. Thus, if $n \leqslant Cd$, then we are done by a brute force search.

Else, $n > Cd$. For each g_i, pick a monomial m_i with degree $\deg(P_i)$. Then there is $i_0 \in [n]$ such that x_{i_0} does not appear in any g_i. Set $g_i' := g_i | x_{i_0} = 0$. Let $\mathcal{B}'$ be the factor defined by the $g_i's$. Note that $\deg(g_i') = \deg(g_i)$ and $\mathrm{depth}(g_i') = \mathrm{depth}(g_i)$. Also, by Lemma 45, $\mathcal{B}'$ is $R_1 - q$-regular.

Now, using recursion, we solve the problem on $n - 1$ variables. That is, decide if for $P' := P | x_{i_0} = 0$ is (k, Δ, Γ)-structured. If P' is not, then P is not either, so we are done. Else, suppose the algorithm does not output NO.

Say

$$P'(x) = \Gamma(S_1(x), \dots, S_k(x)) = \Gamma'(\mathrm{Tr}(\alpha_j S_i(x)) : i \in [k], j \in [r]),$$

where

$$\Gamma'(a_{ij} : i \in [k], j \in [r]) = \Gamma(\sum_j \alpha_i a_{ij} : i \in [k]).$$

Note that while $\Gamma : \mathbb{K}^k \to \mathbb{K}$, we have $\Gamma' : \mathbb{F}^{kr} \to \mathbb{K}$. Let $\mathcal{B}_1$ be the factor formed by $\{\mathrm{Tr}(\alpha_j S_i)\}$. Via the algorithm of [12], regularize $\mathcal{B}' \cup \mathcal{B}_1$ using $R_2 : \mathbb{N} \to \mathbb{N}$ and we get a syntactic refinement $\mathcal{B}' \cup \mathcal{B}_1'$ by the choice of R_1. Let $\mathcal{B}_1' = \{s_1', \dots, s_D'\}$. where

$$\mathrm{Tr}(\alpha_j S_i) = G_{ij}(g_i', s_j' : i \in [C], j \in [D]).$$

Choose R_2 large enough such that the map induced by $\mathcal{B}' \cup \mathcal{B}'_1$ is surjective. Now, fix any $\ell \in [r]$. Then,

$$\mathrm{Tr}(\alpha_\ell P') = G_\ell(g'_1, \ldots, g'_C) = F_\ell(G_{ij}(g'_i, s'_j)),$$

where $F_\ell = \mathrm{Tr}(\alpha_\ell \Gamma')$. Thus, for $a_1, \ldots, a_C, b_1, \ldots b_D \in \mathbb{F}$,

$$G_\ell(a_1, \ldots, a_C) = F_\ell(G_{ij}(a_1, \ldots, b_D) : i \in [C], j \in [D]).$$

Substituting, $a_i = g_i(x)$ and $b_j = 0$ we have

$$\mathrm{Tr}(\alpha_\ell P) = G_\ell(g_1, \ldots, g_C) = F_\ell(G_{ij}(g_i, 0)).$$

Now,

$$\mathrm{Tr}(\alpha_\ell P) = \mathrm{Tr}(\alpha_\ell \Gamma(Q_i : i \in [k])),$$

where $Q_i(x) = \sum_{j=1}^{r} \alpha_j G_{ij}(g'_i, \ldots, 0)$.

Since, this is true for all $\ell \in [r]$, we have

$$P(x) = \Gamma(Q_1(x), \ldots, Q_k(x)).$$

where Q_i is defined as above. This finishes the proof. ◀

References

1. Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Reed-Muller codes. *IEEE Trans. Inform. Theory*, 51(11):4032–4039, 2005. doi:10.1109/TIT.2005.856958.
2. S. Arora and M. Sudan. Improved low-degree testing and its applications. *Combinatorica*, 23(3):365–426, 2003.
3. László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In *Proc. 23rd Annual ACM Symposium on the Theory of Computing*, pages 21–32, New York, 1991. ACM Press.
4. László Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1(1):3–40, 1991.
5. Eli Ben-Sasson, Ghid Maatouk, Amir Shpilka, and Madhu Sudan. Symmetric LDPC codes are not necessarily locally testable. In *Proc. 26th Annual Conference on Computational Complexity (CCC)*, pages 55–65. IEEE, 2011.
6. Arnab Bhattacharyya. Polynomial decompositions in polynomial time. In *Proc. 22nd Annual European Symposium on Algorithms*, pages 125–136, 2014.
7. Arnab Bhattacharyya and Abhishek Bhowmick. Using higher-order fourier analysis over general fields. *CoRR*, abs/1505.00619, 2015. URL: http://arxiv.org/abs/1505.00619.
8. Arnab Bhattacharyya, Victor Chen, Madhu Sudan, and Ning Xie. Testing linear-invariant non-linear properties. *Theory Comput.*, 7(1):75–99, 2011.
9. Arnab Bhattacharyya, Eldar Fischer, Hamed Hatami, Pooya Hatami, and Shachar Lovett. Every locally characterized affine-invariant property is testable. In *Proc. 45th Annual ACM Symposium on the Theory of Computing*, pages 429–436, 2013.
10. Arnab Bhattacharyya, Eldar Fischer, and Shachar Lovett. Testing low complexity affine-invariant properties. In *Proc. 24th ACM-SIAM Symposium on Discrete Algorithms*, pages 1337–1355, 2013.
11. Arnab Bhattacharyya, Elena Grigorescu, and Asaf Shapira. A unified framework for testing linear-invariant properties. In *Proc. 51st Annual IEEE Symposium on Foundations of Computer Science*, pages 478–487, 2010.

12 Arnab Bhattacharyya, Pooya Hatami, and Madhur Tulsiani. Algorithmic regularity for polynomials and applications. In *Proc. 26th ACM-SIAM Symposium on Discrete Algorithms*, pages 1870–1889, 2015.

13 Abhishek Bhowmick and Shachar Lovett. Bias vs structure of polynomials in large fields, and applications in effective algebraic geometry and coding theory. *CoRR*, abs/1506.02047, 2015. URL: `http://arxiv.org/abs/1506.02047`.

14 Abhishek Bhowmick and Shachar Lovett. List decoding Reed-Muller codes over small fields. In *Proc. 47th Annual ACM Symposium on the Theory of Computing*, pages 277–285, New York, NY, USA, 2015. ACM.

15 Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comp. Sys. Sci.*, 47:549–595, 1993. Earlier version in STOC'90.

16 A. Bogdanov and E. Viola. Pseudorandom bits for polynomials. In *Proc.* 48^{th} *IEEE Symp. on Foundations of Computer Science (FOCS'07)*, 2007.

17 Pierre Deligne. Application de la formule des traces aux sommes trigonometriques. In *SGA* $4\frac{1}{2}$ *Springer Lecture Notes in Matematics*, volume 569. Springer, 1978.

18 P. Elias. List decoding for noisy channels. Technical Report 335, Research Laboratory of Electronics, MIT, 1957.

19 Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *J. ACM*, 43(2):268–292, 1996.

20 O. Goldreich and L. Levin. A hard-core predicate for all one-way functions. In *Proc.* 21^{st} *ACM Symposium on the Theory of Computing*, pages 25–32, 1989.

21 O. Goldreich, R. Rubinfeld, and M. Sudan. Learning polynomials with queries: The highly noisy case. *SIAM J. Discrete Math.*, 13(4):535–570, 2000.

22 Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45:653–750, 1998.

23 Oded Goldreich and Tali Kaufman. Proximity oblivious testing and the role of invariances. In *Studies in Complexity and Cryptography*, pages 173–190. Springer, 2011.

24 Oded Goldreich and Dana Ron. On proximity oblivious testing. *SIAM J. Comput.*, 40(2):534–566, 2011.

25 P. Gopalan. A Fourier-analytic approach to Reed-Muller decoding. In *Proc.* 51^{st} *IEEE Symp. on Foundations of Computer Science (FOCS'10)*, pages 685–694, 2010.

26 P. Gopalan, A. Klivans, and D. Zuckerman. List decoding Reed-Muller codes over small fields. In *Proc.* 40^{th} *ACM Symposium on the Theory of Computing (STOC'08)*, pages 265–274, 2008.

27 Parikshit Gopalan, Ryan O'Donnell, Rocco A. Servedio, Amir Shpilka, and Karl Wimmer. Testing Fourier dimensionality and sparsity. In *Proc. 36th Annual International Conference on Automata, Languages, and Programming*, pages 500–512, 2009.

28 William T. Gowers. A new proof of Szeméredi's theorem for arithmetic progressions of length four. *Geom. Funct. Anal.*, 8(3):529–551, 1998.

29 William T. Gowers. A new proof of Szeméredi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

30 Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. *Contrib. Discrete Math.*, 4(2), 2009.

31 Elena Grigorescu, Tali Kaufman, and Madhu Sudan. Succinct representation of codes with applications to testing. *SIAM Journal on Discrete Mathematics*, 26(4):1618–1634, 2012.

32 Alan Guo, Swastik Kopparty, and Madhu Sudan. New affine-invariant codes from lifting. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 529–540. ACM, 2013.

33 V. Guruswami. *List Decoding of Error-Correcting Codes*, volume 3282 of *Lecture Notes in Computer Science*. Springer, 2004.

34 V. Guruswami. *Algorithmic Results in List Decoding*, volume 2 of *Foundations and Trends in Theoretical Computer Science*. Now Publishers, 2006.

35 Hamed Hatami and Shachar Lovett. Estimating the distance from testable affine-invariant properties. In *Proc. 54th Annual IEEE Symposium on Foundations of Computer Science*, pages 237–242. IEEE, 2013.

36 Tali Kaufman and Simon Litsyn. Almost orthogonal linear codes are locally testable. In *Proc. 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 317–326. IEEE, 2005.

37 Tali Kaufman and Shachar Lovett. Worst case to average case reductions for polynomials. In *Proc. 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 166–175, 2008.

38 Tali Kaufman and Dana Ron. Testing polynomials over general fields. *SIAM J. on Comput.*, 36(3):779–802, 2006.

39 Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In *Proc. 40th Annual ACM Symposium on the Theory of Computing*, pages 403–412, 2008.

40 Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky. Inverse conjecture for the Gowers norm is false. In *Proc. 40th Annual ACM Symposium on the Theory of Computing*, pages 547–556, New York, NY, USA, 2008. ACM.

41 R. Pellikaan and X. Wu. List decoding of q-ary Reed-Muller codes. *IEEE Transactions on Information Theory*, 50(4):679–682, 2004.

42 Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. on Comput.*, 25:252–271, 1996.

43 M. Sudan. Decoding of Reed-Solomon codes beyond the error-correction bound. *Journal of Complexity*, 13(1):180–193, 1997. URL: `citeseer.ist.psu.edu/sudan97decoding.html`.

44 M. Sudan. List decoding: Algorithms and applications. *SIGACT News*, 31(1):16–27, 2000.

45 M. Sudan, L. Trevisan, and S. P. Vadhan. Pseudorandom generators without the XOR lemma. *J. Comput. Syst. Sci.*, 62(2):236–266, 2001.

46 Terence Tao. *Higher Order Fourier Analysis*, volume 142 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012.

47 Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields via the correspondence principle. *Analysis & PDE*, 3(1):1–20, 2010.

48 Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. *Ann. Comb.*, 16(1):121–188, 2012.

49 Andre Weil. Sur les courbes algébriques et les varietes qui s'en deduisent. *Actualites Sci. et Ind.*, 1041, 1948.

50 J. Wozencraft. List decoding. Technical Report 48:90-95, Quarterly Progress Report, Research Laboratory of Electronics, MIT, 1958.

51 Yuichi Yoshida. A characterization of locally testable affine-invariant properties via decomposition theorems. In *Proc. 46th Annual ACM Symposium on the Theory of Computing*, pages 154–163, 2014.